

# A Multi-Model Approach for Uncertainty Propagation and Model Calibration in CFD Applications

Jian-Xun Wang, Christopher J. Roy, Heng Xiao\*

*Department of Aerospace and Ocean Engineering, Virginia Tech, Blacksburg, VA 24060, United States*

---

## Abstract

Proper quantification and propagation of uncertainties in computational simulations are of critical importance. This issue is especially challenging for computational fluid dynamics (CFD) applications. A particular obstacle for uncertainty quantifications in CFD problems is the large model discrepancies associated with the CFD models (e.g., Reynolds averaged Navier–Stokes equations) used for uncertainty propagation. Neglecting or improperly representing the model discrepancies leads to inaccurate and distorted uncertainty distribution for the Quantities of Interest. High-fidelity models, being accurate yet expensive, can accommodate only a small ensemble of simulations and thus lead to large interpolation errors and/or sampling errors; low-fidelity models can propagate a large ensemble, but can introduce large modeling errors. In this work, we propose a multi-model strategy to account for the influences of model discrepancies in uncertainty propagation and to reduce their impact on the predictions. Specifically, we take advantage of CFD models of multiple fidelities to estimate the model discrepancies associated with the lower-fidelity model in the parameter space. A Gaussian process is adopted to construct the model discrepancy function, and a Bayesian approach is used to infer the discrepancies and corresponding uncertainties in the regions of the parameter space where the high-fidelity simulations are not performed. The proposed multi-model strategy combines information from models with different fidelities and computational costs, and is of particular relevance for CFD applications, where a hierarchy of models with a wide range of complexities exists. Several examples of relevance to CFD applications are performed to demonstrate the merits of the proposed strategy. Simulation results suggest that, by combining low- and high-fidelity models, the proposed approach produces better results than what either model can achieve individually.

---

\*Corresponding author. Tel: +1 540 231 0926  
*Email address:* [hengxiao@vt.edu](mailto:hengxiao@vt.edu) (Heng Xiao)

## 1. Introduction

Computational simulations, particularly those based on numerical solutions of Partial Differential Equations, have become an indispensable tool for decision-makers by providing predictions of system responses with quantified uncertainties. It is important yet challenging to quantify these uncertainties for the models to provide reliable guidance for decision-making processes.

### 1.1. Propagation of Input Uncertainty in Complex Physical Systems

When a physical system is simulated by using a model, errors may result from various sources, which can be broadly classified into data errors and model errors [1]. Data errors are due to intrinsic variations of the system itself or inexact characterizations thereof, including operation scenarios, initial and boundary conditions, geometry, and material properties, among others. As such, they are also known as *model input uncertainties* or *parametric uncertainties*. Model error or model discrepancy, on the other hand, refers to the differences between the mathematical model and the physical reality, and is often called “mathematical modeling error”, which is the main focus of the present work. Other errors associated with the numerical model including parameter calibration errors and discretization errors are not of particular concern here and are not considered. All the uncertainties propagate into the final predictions, causing difficulties in uncertainty quantification by distorting the output uncertainty distribution. It is important to quantify these uncertainties for the models to provide reliable guidance for decision-making processes. An overview given by Oberkampf et al. [2] highlighted these often under-appreciated challenges. A comprehensive framework has been developed by Roy and Oberkampf [3] for quantifying all major sources of uncertainties in the context of PDE-based numerical simulations.

Interpolation errors are another type of error in uncertainty propagations that result not directly from the numerical models *per se*, but from the limited number of simulations that are conducted by these models. In these cases, simulations used in uncertainty propagation are too expensive to allow for a statistically meaningful number of samples to be propagated. Hence, a small number of simulations are first performed, and then the results are used to build a surrogate model or a response surface. Subsequently, a large number of samples can be propagated by evaluating the surrogate model, which incurs only negligible computational

costs compared to the original model evaluations. The discrepancy between the original expensive model output and the relatively cheap surrogate model output is referred to as interpolation error.

Throughout this paper we use the flow past an airfoil as example, which is a classical problem in Computational Fluid Dynamics (CFD). Despite its simplicity, this prototype problem is of critical importance for fixed-wing aircraft, rotorcraft, pumps, marine and aircraft propellers, and various (gas, wind, tidal and hydraulic) turbines. In this example, data errors include (1) operation conditions such as angle of attack (AoA), Reynolds number, and Mach number, (2) initial and boundary conditions of the flow field, (3) geometry of the airfoil, and (4) material properties (e.g., roughness of the airfoil surface), among others. Characterizing input uncertainties is a delicate issue and a research field by itself [see e.g., 4], but it is beyond the scope of this work. An important assumption is that all input uncertainties are purely probabilistic and characterized with precise probability distributions. Imprecise probabilities, in contrast, are much more challenging to handle [5], and are not pursued in this work. We further assume that the input uncertainties have already been adequately characterized and quantitatively represented with known probability distributions. The quantified input uncertainties are then propagated into predictions of the quantities of interest (QoI) through the complex dynamics of the physical system (e.g., turbulent flow in the example above), leading to uncertainties in the output QoI. Typical QoIs may be integral quantities such as lift and drag coefficients or more detailed field information such as pressure along the airfoil surface and velocity distributions in the wake.

Techniques to propagate uncertainties can be classified into two categories: spectral methods [1] and Monte Carlo (MC) methods [6]. The term “MC methods” in this paper shall refer to all *non-deterministic sampling-based* techniques, without assuming any particular sampling method, e.g., random sampling or Latin hypercube sampling [7]. Spectral methods exploit the global structure of the solution space to enable a mode-based representation. Hence, their effectiveness depends on the smoothness of the prior uncertainties and the input–output response surface [8]. In MC methods each obtained solution carries local information of the solution space, and thus non-smooth response surfaces do not pose a special challenge, although the number of samples required to obtain accurate solution may be increased moderately. Consequently, it is more suitable for turbulent flow problems [9], and thus we will focus on MC methods in this work. However, the framework proposed in this work is equally applicable to other methods (e.g., non-intrusive polynomial chaos [8]).

In MC methods, with a given representation of input uncertainties, the uncertainty quan-

tification process consists of the following major steps: (1) representation of input uncertainty via a collection of samples (referred to as “ensemble”) in the input parameter space, (2) propagation of the ensemble via model evaluations, and (3) reconstruction of the output uncertainties from the propagated ensemble. The second step, i.e., the propagation of the input ensemble, is of particular concern to the present work. If a numerical model that perfectly represents the physical system under investigation were available and computationally affordable (even for a large number of evaluations), this step would be trivial. However, perfect models are rarely available or affordable for any nontrivial physical systems such as turbulent flows, and the discrepancies between model predictions and physical system responses are often appreciable. These model discrepancies distort the output uncertainties and complicate the uncertainty propagation processes. High-fidelity models of turbulent flows, being accurate (if used properly) yet expensive, can accommodate only a small ensemble of simulations and thus lead to large interpolation errors and/or sampling errors; low-fidelity models can propagate a large ensemble, but can introduce large modeling errors (e.g., when Reynolds-Averaged Navier–Stokes (RANS) model is used on flows with massive separation or strong pressure gradient). Therefore, multiple models with different fidelities are employed in this work in a complementary way to address this difficulty. Admittedly, high-fidelity models such as Large Eddy Simulations (LES) or Direct Numerical Simulations are not *guaranteed* to yield more accurate solutions than their low-fidelity counterparts, since they require much more information (e.g., boundary and/or initial conditions), which are often difficult to obtain or specify accurately; on the other hand, low-fidelity models such as RANS can be very accurate when used on the cases they are applicable to, e.g., attached zero-pressure-gradient boundary layers. These cases are not particularly interesting from the perspective of uncertainty propagation, since one can simply use the low-fidelity models to perform a large number of simulations to obtain the QoI uncertainty distribution. Therefore, in this work we are mainly concerned with cases that are challenging for the cheap, low-fidelity model but amiable for the expensive, high-fidelity model (e.g., free-shear flows, for which LES often have improved predictions over RANS).

### *1.2. Multi-Model Strategies for Uncertainty Propagation*

The benefits of utilizing computational models of multiple fidelities have long been recognized by the science, engineering, and statistics communities. Various statistical frameworks have been proposed to incorporate information from multiple models. Here we review several categories of methods that are most closely related to the present work.

Bayesian Model Averaging (BMA) method [10] and its dynamic variant [11] compute the distributions of the prediction by averaging the posterior distribution under each of the models considered, weighted by their posterior model probability. However, the fundamental assumption in BMA is that the models of concern are all plausible, competitive explanations or fitting of the observed data. There is no hierarchy among these models. While this context may be relevant for certain applications (e.g., correlation of health factors related to a disease), in engineering we often have a well-defined hierarchy of models from cheap, low-fidelity models to expensive, high-fidelity models, and thus BMA methods are not of particular interest in the context of uncertainty quantification in CFD concerned in this study.

In contrast to BMA methods, for which a model hierarchy does not exist, Multi-Level Monte Carlo (MLMC) methods create and utilize such a hierarchy to facilitate uncertainty propagation by running the same simulator on a series of successively refined meshes, leading to a sequence of multi-level models. By optimizing the number of simulations conducted on each level, significant speedup compared to the standard MC method has been obtained [12, 13]. MLMC methods have been used successfully on stochastic PDE-based path simulations in financial engineering [14], on ground-water flows [12], and more recently on inviscid compressible flows [15]. A recent extension of MLMC by Mueller et al. [16] from multiple meshes to multiple models is noteworthy. However, essentially the MLMC method is a control variate based variance reduction technique with two critical assumptions [14]: (1) the ensembles computed on different levels are all *unbiased* estimators of the QoI, and (2) the variance of the estimator decreases as the mesh is refined. These assumptions are also applicable to the multi-model extension [16]. Therefore, model discrepancies are not accounted for in MLMC methods. In turbulent flows applications, however, model discrepancies play a pivotal role, and thus the MLMC method is likely to be not applicable, since the assumptions above cannot be fully justified.

Data assimilation represents another approach of combining multiple sources of information. It is a method widely used in geosciences, particularly in the operational weather forecasting community for assimilating numerical model predictions and observation data. Recently, Narayan et al. [17] proposed a framework to assimilate multiple models and observations, which expresses the assimilated state as a general linear functional of the models and the data, allowing different sections of the state vector to have different weights. In general, model discrepancies are not explicitly considered in data assimilation methods, although the available data may implicitly correct any biases in the model predictions in the filtering

operations. Similar to MLMC methods, the filtering operations in data assimilations aim to reduce variances of the overall estimation from several unbiased sources. Narayan et al. [18] proposed a multi-fidelity approach for stochastic collocation methods, and Zhu et al. [19] tested the method on non-trivial PDEs.

Gaussian process (GP) modeling is a promising approach to account for model discrepancies. The basic assumption is that the quantities of interest at different design points obey multivariate joint Gaussian distributions. With this assumed correlation, Bayesian updating can be used to infer the distribution of the QoI in the entire field from the data available at certain locations (in the parameter space or the physical space, depending on the application). Results obtained from the Bayesian inference can be viewed as a probabilistic response surface, i.e., mean predictions with quantified uncertainties or confidence intervals (CI). Often referred to as “kriging”, Gaussian processes are used in geostatistics applications to reconstruct permeability fields from drilling data at a limited number of locations, which are expensive to obtain and thus are often sparse [20]. In engineering design optimization it is used to build surrogate models based on simulations performed at a small number of points in the parameter space [21].

Kennedy and O’Hagan [22] developed the first multi-model prediction framework accounting for model discrepancies by using Gaussian process and Bayesian inference. An important assumption of their framework is that the predictions from different levels are auto-regressive. That is, the higher-level model prediction  $z^H$  is related to the lower level model prediction  $z^L$  by a regressive coefficient  $\rho$  and a discrepancy  $\delta$ , i.e.,  $z^L = \rho z^H + \delta$ , where  $\delta$  is described by a GP. Another assumption is that the priors for the prediction of each model can be described by stationary GP. Simulation data obtained from all models are then used to infer the predictions of the highest-fidelity model. More than a decade after being proposed, this elegant and powerful framework has yet to find widespread use in the engineering community, particularly in CFD applications. Recent applications in the statistics community include Refs. [23] and [24]. A major obstacle for its application in engineering systems with complex physics is the GP assumption on priors of model predictions. As response surfaces in engineering and particularly in fluid dynamics are often highly complex and regime-dependent, this assumption should be revisited.

### *1.3. Objective and Summary of This Work*

The objective of this work is to account for model discrepancy in uncertainty quantification for CFD applications based on information from models of various fidelities.

Taking the airfoil problem again for example, one is often interested in the relation between the lift coefficient (i.e., normalized lift force) and the angle of attack (AoA, defined as the angle between the airfoil chord and the incoming flow). The prediction can be obtained by a low-fidelity model such as the panel method based on potential flow assumption [25], a medium-fidelity model such as Reynolds-Averaged Navier–Stokes (RANS) equation solver [26], and a high-fidelity model such as Large Eddy Simulation (LES) [27]. The above-mentioned devices such as airfoils and turbine blades normally operate at small AoA with attached flow, a regime where all three models predict the lift quite well. As the AoA increases, mild flow separation occurs. In this regime the low-fidelity model (panel method) fails to capture majority of the flow physics. While the performance of the medium fidelity model (RANS) also deteriorates, it captures the general trend well. At very high AoA (approximately  $12^\circ$ – $16^\circ$  for a typical airfoil [28]), massive flow separation occurs, leading to dramatically decreased lift and sharply increased drag, a regime referred to as “stall” [29]. This condition is to be avoided in most operations, but may occur at off-design conditions or extreme operation conditions (e.g., fighter jet maneuvering). Only the highest fidelity model (LES) can give reliable predictions in this regime. The model discrepancy of a RANS solver would have general trends as shown in Fig. 1, which shows good prediction skill for attached flows, slightly deteriorated performance for mildly separated flows, and catastrophic failure near stall. The three regimes are demarcated in Fig. 1. Model discrepancies for other models would be different but have similar regime-dependent features.

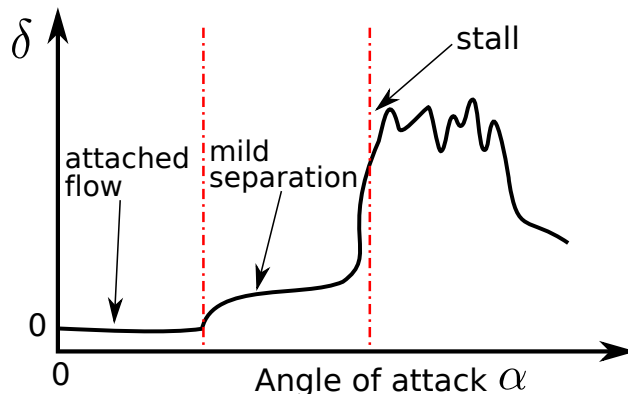


Figure 1: A schematic representation of the model discrepancy  $\delta$  for a RANS model at different angle of attack  $\alpha$ , highlighting the regime-dependence feature of  $\delta$ . The three different flow regimes are demarcated with dash–dotted lines.

From the example above, we make a few general observations about the specific features

of uncertainty propagation in CFD application:

1. Stationary Gaussian processes are likely to be inadequate in describing the model discrepancies, and non-stationary GPs are more appropriate. However, non-stationary GPs have more degrees of freedom (e.g., parameters), and inferences of these parameters would be challenging, particularly when the available data are sparse (i.e., when very few high-fidelity simulations can be afforded).
2. The low-fidelity models (e.g., panel methods or even empirical models) are relatively cheap computationally, with each evaluation taking only seconds or minutes. Consequently, the input parameter space can be sampled sufficiently when performing low-fidelity simulations, and thus Gaussian process modeling of the interpolation errors for the low-fidelity predictions as conducted in Kennedy and O’Hagan [22] is largely unnecessary.
3. Correlations among predictions given by models of different types (e.g., panel method, RANS, and LES) are weak and can be regime-dependent. This is in stark contrast to the scenarios assumed in Kennedy and O’Hagan [22], where different model levels correspond to the same code running at different mesh resolutions (as in MLMC methods).

In light of these observations, we propose a simpler method based on the classical Bayesian approach. Specifically, we describe only the model discrepancy  $\delta$  between the low- and high-fidelity models ( $z^L$  and  $z^H$ , respectively) with Gaussian processes without assuming the responses themselves to have a certain distribution. By running low- and high-fidelity simulations on a small number of selected samples in a small ensemble  $\mathcal{D}^H$ , a Gaussian process  $\mathcal{GP}$  is fitted for the discrepancy  $\delta$ . We propagate a large ensemble  $\mathcal{D}^L$  by using a low-fidelity model  $z^L$ , and then use realizations of model discrepancies (i.e., random draws from the fitted Gaussian process  $\mathcal{GP}$ ) to correct  $z^L$ . With this strategy the distortion caused by low-fidelity model deficiency on the output uncertainty distribution can be reduced with the information provided by the high-fidelity simulations. In addition to the obvious yet essential fact that Kennedy and O’Hagan’s work [22] focused on prediction while our work is concerned with uncertainty propagation, the most notable differences between this work and their approach of utilizing multi-model predictions are as follow:

1. No assumptions are made on the model predictions themselves. Hence, prior beliefs on model predictions are not needed. Interpolation errors on the low-fidelity model level



are thus avoided.

2. The auto-regressive relation among the models is not assumed. Consequently, no attempts are made to infer the regression coefficient, and the high-fidelity simulations are used only to fit the Gaussian process for the model discrepancy. This is a more efficient use of data from the expensive, high-fidelity model  $z^H$ .
3. Instead of inferring the output of the highest model level conditional on all levels of simulations as in [22], we propagate MC samples directly with the low-fidelity model  $z^L$ , and then use the inferred model discrepancies to correct the output ensemble.

In this work we perform a proof-of-concept investigation on the feasibility of the multi-model method for CFD applications by using two model levels, i.e., a low-fidelity model  $z^L$  and a high-fidelity model  $z^H$ . A stationary GP is used to fit the simulation data of model discrepancy  $\delta = z^H - z^L$ . Future extensions will be made by using more than two model levels and by modeling the discrepancy  $\delta$  with non-stationary Gaussian processes.

Previous authors have used GPs and multi-model framework to improve predictions [e.g., 23, 24] and to serve as surrogate models in design optimizations [21, 30]. However, to the authors' knowledge, the use of GPs and multi-fidelity models to account for model discrepancies and to perform implicit model calibrations in uncertainty propagation is a relatively novel development, particularly in a context relevant to complex physical system such as those in CFD applications. A unique issue of using GP-based probabilistic response surface for uncertainty propagation that is not present in using it for predictions or optimizations is the coupling between the model uncertainties and the input uncertainties. This issue will be discussed in the paper (see Section 4.2).

The rest of the paper is organized as follows. The methodology and algorithm of the proposed multi-model uncertainty propagation scheme based on Gaussian process modeling of discrepancies are presented in Section 2. Four test cases are investigated to demonstrate the proposed method, and the results are presented in Section 3. The comparison of the multi-model strategy with traditional single-model approaches, as well as a general reflection of the proposed method is discussed in Section 4. Finally, Section 5 concludes the paper.

## 2. Methodology

An essential part of the proposed method is to use the high-fidelity simulations performed on a small ensemble (along with the low-fidelity simulations on the same ensemble) to construct a probabilistic response surface for the model discrepancy  $\delta$ . Random draws from the

probabilistic response surface are then used to correct the low-fidelity simulations, which are performed on an ensemble as large as required. The correction of the low-fidelity results by using data from the high-fidelity simulations can be considered as an implicit form of model calibration. In this section, we first introduce Gaussian process modeling and Bayesian inference of model discrepancy in Section 2.1. The overall algorithm of the multi-model uncertainty propagation and model calibration method is then presented in Section 2.2.

### 2.1. Gaussian Process Modeling and Bayesian Inference of Model Discrepancy

Referring to the airfoil example again, the model discrepancy  $\delta$  is approximated by the difference between the lift coefficients predicted by the low- and high-fidelity models, and its magnitude depends on the AoA, denoted as input variable  $x$ . Since it is prohibitively expensive to perform high-fidelity simulations of the flow over an airfoil for a large number of AoA, the discrepancy  $\delta$  is effectively an unknown function of  $x$ . To facilitate modeling, we assume (1) that the discrepancy  $\delta(x_i)$  at any location  $x_i$  is a random variable obeying a Gaussian distribution and (2) that the discrepancies  $\delta(x_i)$ , where  $i = 1, \dots, n$ , at any number of  $n$  locations have a joint Gaussian distribution. Consequently, the unknown function  $\delta(x)$  is a Gaussian process, which is denoted as [31]:

$$\delta(x) \sim \mathcal{GP}(m, k), \quad (1)$$

where the mean function  $m(x)$  is the expectation of  $\delta(x)$ ; the covariance or kernel function  $k(x, x')$  dictates the covariance between the values of function  $\delta$  at two locations  $x$  and  $x'$ .

In this work it is assumed that the low-fidelity model is explicitly calibrated beforehand with best available information (but not against high-fidelity simulation results) to eliminate any bias over the entire input parameter domain, and thus a zero mean function  $m(x) = 0$  is used. Assuming  $\delta(x)$  is a smooth function over most part of the parameter domain, the following kernel function is chosen, i.e.,

$$k(x, x') = \sigma_f^2 \exp\left(\frac{-|x - x'|^2}{2l^2}\right) \quad (2)$$

where  $|\cdot|$  denotes Euclidean norm,  $\sigma_f$  determines the overall magnitude of the variance, and  $l$  is the length scale.  $\sigma_f$  and  $l$  are referred to as hyperparameters. This kernel function is stationary in that the covariance between two points only depends on their distances and not on their specific locations, i.e.,  $k(x - x') = k(|x - x'|)$ . While this may not be a good assumption in some cases as pointed out in Section 1.3, it can be a reasonable choice when only a small amount of data is available, and it is thus adopted in the proof-of-concept study.

More sophisticated kernel functions are possible [31–34], but this issue is beyond the scope of this work. Investigations of using non-stationary kernel functions are in progress and will be published separately.

If at certain locations  $\mathbf{x}_o = [x_1^o, \dots, x_{n_{obs}}^o]^T$  in the parameter space the values of the discrepancy are found to be  $\mathbf{y}_o = [\delta_1^o, \dots, \delta_{n_{obs}}^o]$ , the objective is to predict the values  $\mathbf{y}_p$  of the function  $\delta$  at some other locations  $\mathbf{x}_p = [x_1^p, \dots, x_{n_{pred}}^p]^T$ . Subscripts  $o$  and  $p$  indicate “observations” and “predictions”, respectively. Observations in this work refer the process of performing low- and high-fidelity simulations to “observe” the values of the discrepancy function at locations  $\mathbf{x}_o$ .

Following Arendt et al. [35], a modular approach is adopted here, i.e., the hyperparameters are first inferred from observation  $\mathbf{y}_o$  via Maximum Likelihood Estimation (MLE) [36] before the predictions  $\mathbf{y}_p$  are inferred. Specifically, the hyperparameter pair  $\sigma_f$  and  $l$  that maximizes the likelihood of obtaining observation  $y_o$  is chosen. In practice, the logarithm of the likelihood  $p$  of  $\mathbf{y}_o$  conditioned on hyperparameters, i.e.,

$$\log p(\mathbf{y}_o | \mathbf{x}_o, \sigma_f, l) = -\frac{1}{2} \mathbf{y}_o^T K_o \mathbf{y}_o - \frac{1}{2} \log \det(K_o) - \frac{n_{obs}}{2} \log(2\pi) \quad (3)$$

is maximized, where “|” means conditional upon and  $\det(K_o)$  is the determinant of matrix  $K_o$ . The optimization is performed with standard gradient-based procedure [31]. Once chosen, the hyperparameters are considered fixed in subsequent inference of  $\mathbf{y}_p$ .

Before the values of  $\mathbf{y}_o$  are known, the elements in the stacked vector  $[\mathbf{y}_o, \mathbf{y}_p]^T$  have a joint Gaussian distribution according to the definition of a GP,

$$\begin{bmatrix} \mathbf{y}_o \\ \mathbf{y}_p \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} K_o & K_{op} \\ K_{op}^T & K_p \end{bmatrix} \right), \quad (4)$$

where  $\mathcal{N}$  denotes normal (Gaussian) distribution. The mean vector  $[\mathbf{0}, \mathbf{0}]^T$  is the value of mean function  $m(x)$ , chosen to be zero, evaluated at  $[\mathbf{x}_o, \mathbf{x}_p]^T$ . The covariance matrix  $K$  is obtained in a similar way by evaluating the kernel function  $k$ , and is partitioned to sub-matrices  $K_o$ ,  $K_p$  and  $K_{op}$ , corresponding to the auto-variances of  $y_o$  and  $y_p$ , and their cross-covariance, respectively. When  $\mathbf{y}_o$  is known, the distribution of  $\mathbf{y}_p$  conditional on  $\mathbf{y}_o$  is still joint Gaussian, which can be obtained with standard Bayesian inference procedure as the following [31]:

$$\mathbf{y}_p | \mathbf{y}_o, \sigma_f, l \sim \mathcal{N}(K_{op}^T K_o^{-1} \mathbf{y}_o, K_p - K_{op}^T K_o^{-1} K_{op}), \quad (5)$$

which is called the posterior distribution of  $\mathbf{y}_p$ . In contrast, Eq. (4) specifies the prior distribution of  $\mathbf{y}_p$ . That is, the distribution of  $\mathbf{y}_p$  can be extracted from Eq. (4) as  $\mathbf{y}_p \sim \mathcal{N}(\mathbf{0}, K_p)$ .

These are illustrated in Fig. 2, which will be further discussed later in Section 3.2.1. The prior distribution of  $\delta$ , which is specified as a Gaussian process with a zero mean and a stationary kernel, is shown in Fig. 2a. The posterior distribution of  $\delta$  given observation data at five points, which is still a Gaussian process [31], is shown in Fig. 2b. The mean functions and 95% confidence intervals in both the prior and the posterior distributions are indicated. The changes in the mean and the covariance of the posterior distribution compared to those in the prior reflect the new information provided by the observations  $\mathbf{y}_o$ .

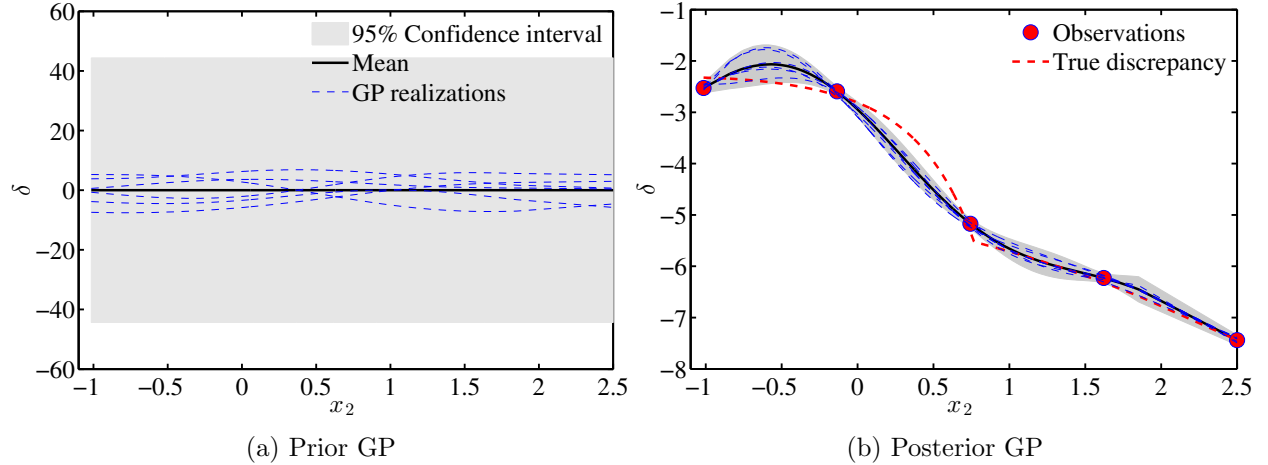


Figure 2: (a) The prior distribution of the model discrepancy function  $\delta$  is represented as a Gaussian process (GP) with a zero mean function and a stationary kernel. (b) The posterior GP of  $\delta$  given observation data at five input locations. In both the prior and the posterior GPs, the mean functions, 95% confidence intervals, and several random realizations from the GPs are indicated.

## 2.2. Multi-Model Uncertainty Quantification Method

After detailing the procedure of constructing a probabilistic response surface for the model discrepancy above, which is at the core of the proposed multi-model method, we outline below the overall algorithm of the uncertainty propagation referring to the airfoil example above.

Given the joint uncertainty distribution of input  $x$ , e.g., the operation conditions (AoA and Reynolds number), the objective is to find the uncertainty distribution of the QoI, e.g., the lift coefficient. Assuming, again, that the input uncertainties have been characterized

with known probabilities, a traditional MC uncertainty propagation procedure consists of the following three steps:

**Sampling:** Sample the input uncertainty with a space filling method (e.g., the Latin Hypercube sampling [7]) to obtain an ensemble  $\mathcal{D}^L = \{x_j\}_{j=1}^{N^L}$ , where  $j = 1, \dots, N^L$ , and  $N^L$  is the number of samples in  $\mathcal{D}^L$ .

The notation  $\{\phi_j\}_{j=1}^n$  indicates a set with  $n$  elements  $\phi_1, \dots, \phi_n$ . This notation is used to indicate sets and ensembles throughout the paper.

**Propagation:** Propagate each sample  $x_j$  in  $\mathcal{D}^L$  with a model  $z^L$ , i.e.,  $f_j = z^L(x_j)$ , leading to an output ensemble  $\mathcal{F}^L = \{f_j\}_{j=1}^{N^L}$ .

**Aggregation:** Aggregate the output ensemble  $\mathcal{F}^L$  to obtain the output uncertainty distribution of the QoI.

If the responses given by the numerical model  $z^L$  differ from the true response, the obtained QoI uncertainty distribution would be distorted. Without introducing further independent information, it is difficult, if not impossible, to quantify the nature of this distortion or to correct it. Increasing the fidelity of the model used in the propagation step above would generally reduce the model discrepancy and thus decrease the distortion in the output uncertainty. However, high-fidelity models have high computational costs and would reduce the number of samples that can be afforded, leading to increased sampling errors. Note that increasing the number  $N^L$  of samples in the input ensemble  $\mathcal{D}$  alone would reduce the sampling error, but this does not influence the modeling error. With given computational resources, increasing the number of samples and increasing the model fidelity contradict each other by competing for resources.

In view of these difficulties, in the proposed method we perform high-fidelity simulations on a small number  $N^H$  of samples in  $\mathcal{D}^H$  to obtain an ensemble  $\mathcal{F}^H$ . Assuming a GP for the prior of the model discrepancy  $\delta$  as in Eq. (4), the simulation results in ensemble  $\mathcal{F}^H$  are then used to obtain the posterior distribution according to Eq. (5). A number  $N_c$  of realizations of the discrepancy function are drawn from the posterior distribution to correct the output ensemble  $\mathcal{F}^L$  given by the low-fidelity model, with each realization corresponding a corrected ensemble. The corrections lead to a set of  $N_c$  ensembles, which is denoted as  $\{\hat{\mathcal{F}}_k^L\}_{k=1}^{N_c}$ . In this case each element in the set is an ensemble  $\mathcal{F}_k^L$ , where  $k = 1, \dots, N_c$ . These ensembles are then aggregated to obtain the output uncertainty for the QoI. The proposed procedure is illustrated with schematic shown in Fig. 3. In comparison,

the modules enclosed in the dashed box indicate the procedures of the traditional MC method as outlined above. As mentioned briefly in the beginning of this section, the multi-model strategy for uncertainty quantification uses data from high-fidelity simulations to construct model discrepancy functions to correct low-fidelity model results. This procedure, which is not present in traditional MC methods with single-model approaches, can be considered as an implicit form of model calibration. This calibration differs from tradition calibration procedures in two major aspects: (1) it takes into account the errors of calibration data themselves (i.e., the high-fidelity simulation), and (2) it is probabilistic and not deterministic, giving a confidence interval on the model discrepancy based on GP assumptions.

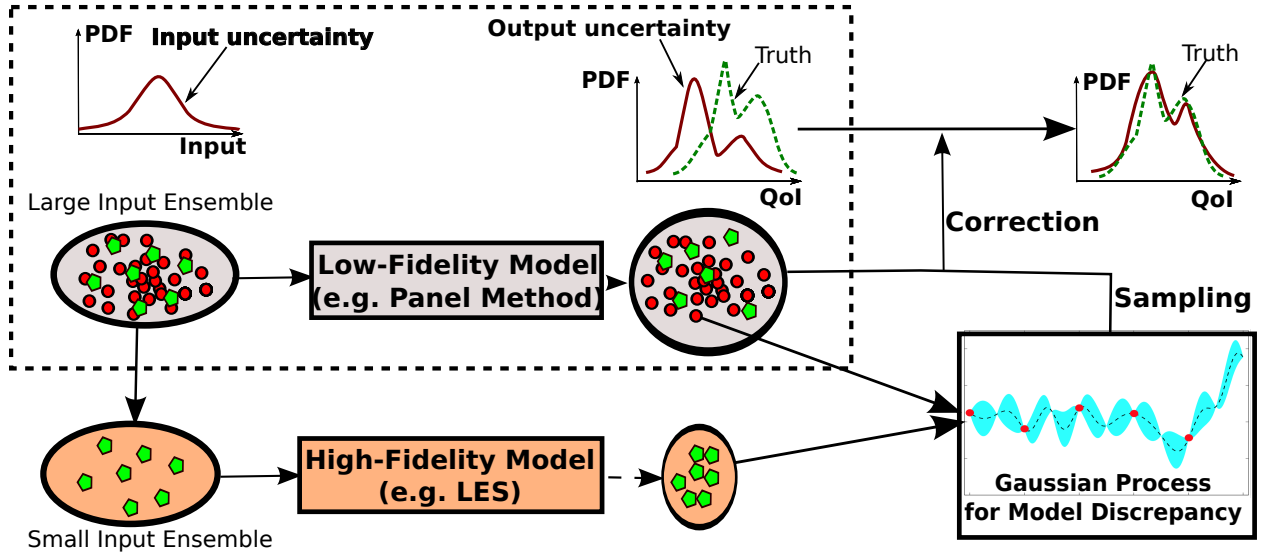


Figure 3: Schematic of the proposed multi-model strategy for uncertainty propagation with implicit model calibration. A computationally cheap, low-fidelity model is used to propagate a large number of samples. Computationally expensive, high-fidelity simulations are used to obtain observations of the model discrepancy, based on which a probabilistic response surface for the discrepancy is constructed by using Gaussian prior assumption and Bayesian inferences. Realizations of the model discrepancy are drawn from the constructed probabilistic response surface to correct the low-fidelity model predictions and to obtain corrected ensembles. This correction procedure is considered as an implicit calibration of the low-fidelity model by using the high-fidelity model results.

The overall algorithm is summarized as follows:

1. **(Sampling)** Generate ensemble  $\mathcal{D}^L = \{x_j^L\}_{j=1}^{N^L}$  from given input uncertainty distribu-

tion.

2. (**Propagation**) Propagate each sample in  $\mathcal{D}^L$  through low-fidelity model  $z^L$  to obtain output ensemble  $\mathcal{F}^L = \{f_i^L\}_{i=1}^{N^L}$ , where  $f_i^L = z^L(x_i^L)$ .
3. Generate input ensemble  $\mathcal{D}^H = \{x_j^H\}_{j=1}^{N^H}$  for discrepancy observation.
4. Propagate ensemble  $\mathcal{D}^H$  with both low- and high-fidelity models  $z^L$  and  $z^H$  to obtain observations for model discrepancy, obtaining ensemble  $\mathcal{F}^H = \{\delta_i\}_{i=1}^{N^H}$  with  $\delta_i = z^H(x_i^H) - z^L(x_i^H)$ .
5. Compute hyperparameters  $\sigma_f$  and  $l$  by maximizing  $\log p$  in Eq. (3).
6. Construct posterior distribution  $\mathcal{GP}(m, k)$  for model discrepancy function  $\delta(x)$  from observations  $\mathcal{F}^H = \{\delta_i\}_{i=1}^{N^H}$  and hyperparameters  $\sigma_f$  and  $l$  according to Eq. (5), and sample  $N_c$  realizations  $\{\delta_k(x)\}_{k=1}^{N_c}$  from the posterior.
7. Correct output ensemble  $\mathcal{F}^L$  by each realization  $\delta_k$  to obtain corrected ensembles  $\hat{\mathcal{F}}_k^L$  where  $k = 1, \dots, N_c$ .
8. (**Aggregation**) Aggregate all corrected ensembles  $\{\hat{\mathcal{F}}_k^L\}_{k=1}^{N_c}$  to obtain the output uncertainty distribution of the QoI.

Steps 1, 2, and 8 correspond to the same steps in the traditional Monte Carlo method described above. Other steps are unique to the multi-model method and are used to account for model discrepancies.

Three sampling procedures are required in the multi-model uncertainty propagation method outlined above. The first one, which is used to obtain design set  $\mathcal{D}^L$  in Step 1, is the same as in traditional MC methods and does not need further discussion. In this work we used Latin hypercube sampling to perform this sampling. Another sampling is required to obtain design set  $\mathcal{D}^H$ . Since points in this ensemble are used to construct Gaussian processes, we use a deterministic, uniform sampling on the range of parameter domain to minimize the largest possible distance between an arbitrarily chosen point in the domain to the nearest sampled points. More sophisticated sampling schemes are possible [37], which will be pursued in future studies. Finally, Step 6 involves sampling a *function* (with infinite number of degrees of freedom) from the prior GP as described in Eq. (5). In practical implementations, only the values of the function at the parameter locations corresponding to the samples in the ensemble  $\mathcal{D}^L$  are needed. Therefore, we only sample  $N^L$  random numbers with joint

Gaussian distribution  $\mathcal{N}(\mathbf{0}, K)$ , where  $K$  is the blocked kernel matrix in Eq. (5). This is performed with standard procedures [31], i.e., a vector  $\tilde{\mathbf{x}}$  consisting of  $N^L$  independent and identically distributed (i.i.d) random numbers is generated first, and then a transformation  $\mathbf{x} = L\tilde{\mathbf{x}}$  is performed to obtain a random vector with desired covariance  $K$ , where  $LL^T = K$ . Eigenvalue decomposition is used to obtain  $L$  from  $K$ , whose computational cost scales as  $O(N^3)$  with  $N$  being the size of the random vector  $\mathbf{x}$ . This can be expensive when  $N$  is large. However, it is worth pointing out that the sample does not involve numerical model evaluations. Various standard techniques are available to reduce the computational expense of this operation, e.g., by partitioning the parameter space via treed schemes [32].

### 3. Numerical Simulations

In this section we evaluate the method proposed in Section 2 on a few synthetic cases that are simple yet of relevance to CFD applications. Four cases with a wide range of complexities are chosen to illustrate the proposed multi-model uncertainty propagation method.

#### 3.1. Test Cases and Parameters

##### 3.1.1. Choice of Synthetic Test Cases

When choosing test cases, we have the typical problem of airfoil lift coefficient prediction in mind. In line with the example given above, one is interested in the uncertainty distribution of the lift coefficient  $C_L$  when the operation conditions (i.e., input variables such as AoA  $\alpha$  and Reynolds number  $Re$ ) conform to known distributions. The uncertainty propagation involves performing numerical simulations on a number of samples from the input  $x = (\alpha, Re)$  space to obtain samples in the output  $C_L$  space. The simulations can be conducted by using a cheap, low-fidelity model such as panel method, or an expensive, high-fidelity model, or any models with computational costs and fidelities in between. The  $C_L(\alpha, Re)$  response surface for a much studied classical airfoil, NACA-0015, is shown in Fig. 4, which are fitted from experimental data [28]. A notable feature of this response surface is the smooth, monotonic increase of  $C_L$  with respect to AoA  $\alpha$  in most regions and a sharp decrease at certain AoA, corresponding to stall. The flow physics and regime changes are dominated by the AoA, but note that the Reynolds number provides modulations as it influences the angle at which the stall occurs.

It is possible to use the response surface as presented in Fig. 4 as the ground truth, and use a numerical models with a hierarchy of fidelities (e.g., panel method and LES) to perform the uncertainty propagation. However, several difficulties exist with this approach. The true



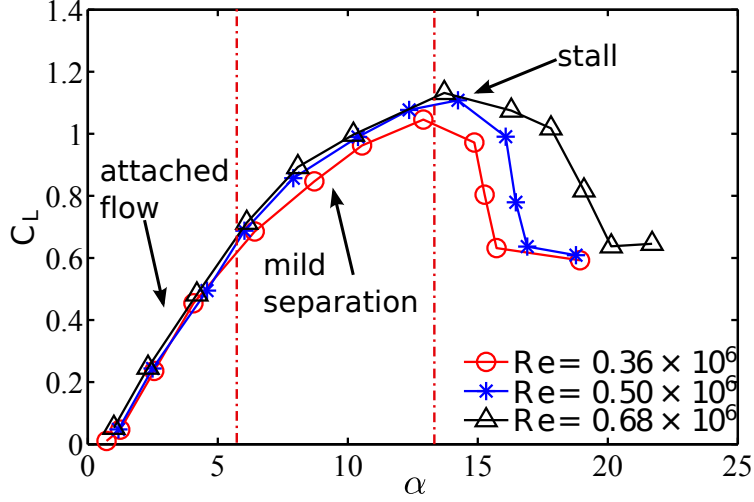


Figure 4: The dependence of the lift coefficient  $C_L$  of a NACA-0015 airfoil section on AoA and Reynolds number  $Re$ . The dash-dotted lines approximately delineate the same three regimes as illustrated in Fig. 1.

response surface is not complete due to the sparseness of the available experimental data. Moreover, high-fidelity simulations are expensive to perform and involve many complexities of their own, which are not essential to the present study of uncertainty propagation. Therefore, for the purpose of evaluating the proposed method, we use a few synthetic response surfaces and numerical models, which closely mirror the features of the actual response surfaces and models in the airfoil example and in other CFD applications.

Two response surfaces with different complexities are used, both of which have two input variables, corresponding to two-dimensional input parameter spaces. They were originally used by Helton and Davis [7] to evaluate and compare the performances of different sampling methods for MC simulations. The first response surface has the following expression:

$$f_1(x_1, x_2) = x_1 + x_2 + x_1x_2 + x_1^2 + x_2^2 + x_1 \min(\exp(3x_2), 10). \quad (6)$$

which is a monotonically increasing function with a change of characteristics near plane  $x_2 = 0.7$ , as plotted in Fig. 5. The specific locations of the characteristics change depend on the value  $x_1$ . In this mapping,  $x_2$  dominates the feature of the response surface, and  $x_1$  provides minor modulations, which is reminiscent of the roles of  $\alpha$  and  $Re$  in the airfoil example. One can compare Fig. 5b with Fig. 4 to appreciate this analogy.

Being a monotonic mapping, the synthetic response surface in Eq. (6) is slightly less challenging for uncertainty quantifications than the actual  $C_L$  response surface. Hence, a

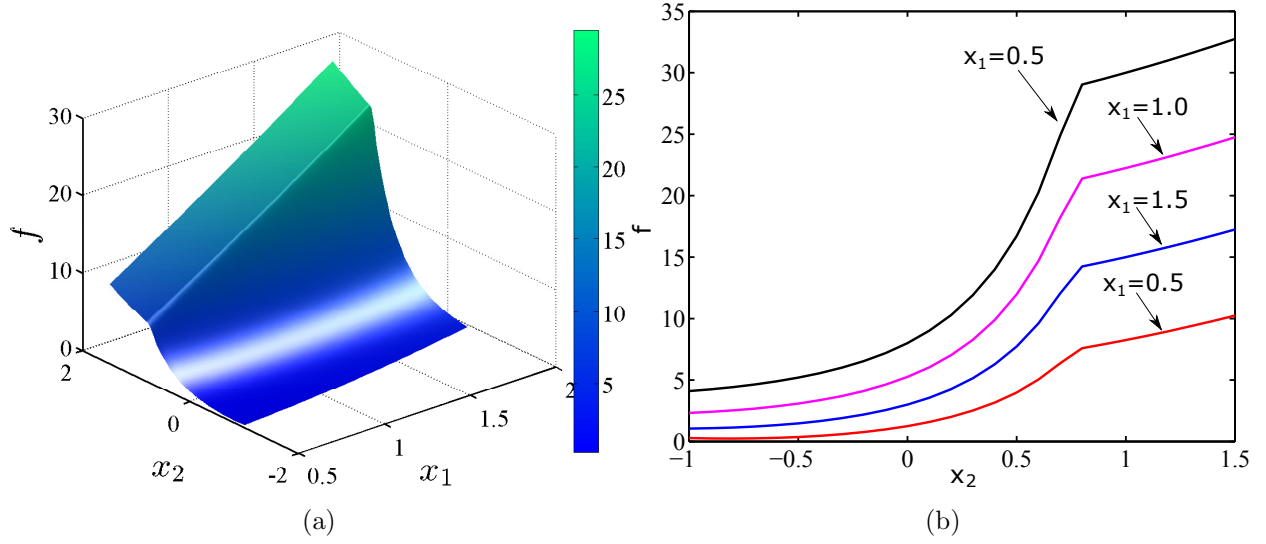


Figure 5: Plots of the monotonic mapping  $f(x_1, x_2)$  in Eq. (6) showing (a) the three-dimensional elevated surface with contour and (b) the cross-sections at several values  $x_1$ .

more challenging response surface with non-monotonic application and high-wave-number oscillations is adopted as well to evaluate the proposed method [7]:

$$f_2(x_1, x_2) = x_1 + x_2 + x_1x_2 + x_1^2 + x_2^2 + x_1g(x_2) \quad (7)$$

$$\text{where } g(x_2) = \begin{cases} 10 & \text{if } h(x_2) \geq 10 \\ h(x_2) & \text{if } |h(x_2)| < 10 \\ -10 & \text{if } h(x_2) \leq -10 \end{cases}$$

$$\text{and } h(x_2) = (x_2 - 11/43)^{-1} + (x_2 - 22/43)^{-1} + (x_2 - 33/43)^{-1}$$

The non-monotonic function is plotted in Fig. 6 in the same way as for the monotonic mapping.

The two cases above both have two input variables. To facilitate graphical presentations and to help understand the basic properties of the proposed method, we first test two one-dimensional response surfaces, which are obtained by setting the less important variable  $x_1$  to a fixed value  $x_1 = 1.25$  in the two mapping functions in Eqs. (6) and (7) above. This simplification reduces the complexity of the response surface while retaining its essential features. The 1D monotonic and non-monotonic response surfaces obtained in this way resemble one of the lines in Figs. 5b and 6b, respectively.

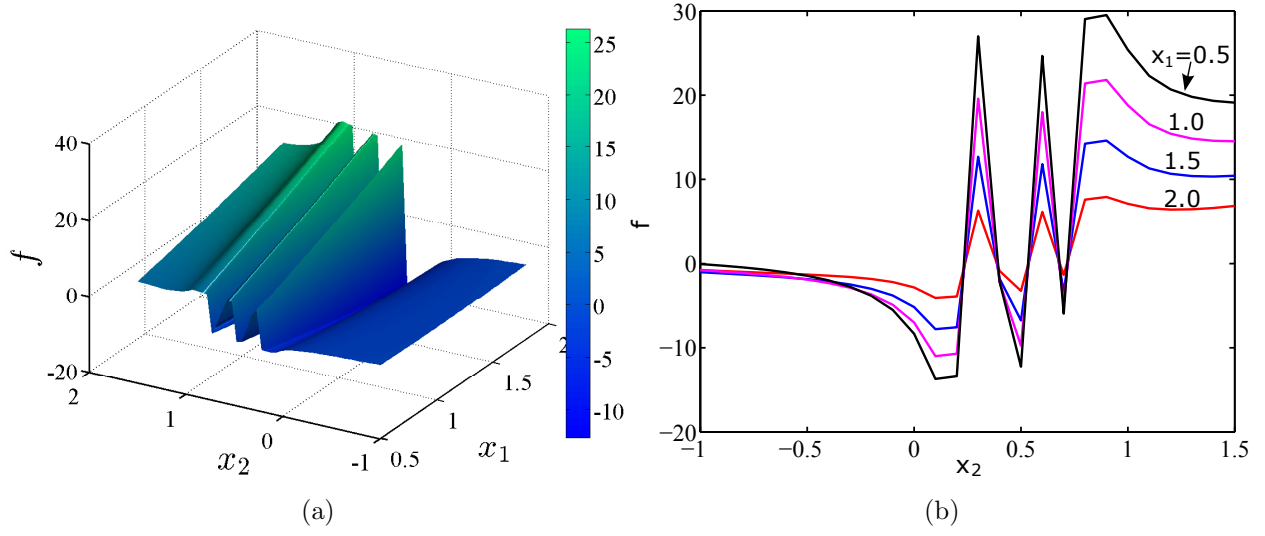


Figure 6: Plots of the non-monotonic mapping  $f(x_1, x_2)$  in Eq. (7) showing (a) the three-dimensional elevated surface with contour and (b) the cross-sections at several values  $x_1$ .

In summary, four synthetic response surfaces of relevance to CFD applications are used in this work to test the proposed uncertainty quantification method:

**Case 1:** Monotonic 1D mapping as in Eq. (6) with  $x_1 = 1.25$ ,

**Case 2:** Non-monotonic 1D mapping as in Eq. (7) with  $x_1 = 1.25$ ,

**Case 3:** Monotonic 2D mapping as in Eq. (6), and

**Case 4:** Non-monotonic 2D mapping as in Eq. (7).

### 3.1.2. Low- and High-Fidelity Models

To mimic the behaviors of low-fidelity models (e.g., panel method or RANS solvers) and high-fidelity models (e.g., LES) in CFD applications, the low-fidelity model used in this proof-of-concept study is constructed by adding a discrepancy function  $\delta(x)$  to the true response  $f(x)$ , with the discrepancy consisting of a constant  $B$  and a part that is proportional to  $f$ , i.e.,

$$z^L(x) = f(x) + \delta(x) \quad \text{with} \quad \delta(x) = \nu f(x) + B \quad (8)$$

where  $\nu$  is the proportionality constant. Larger values of  $\nu$  and  $B$  correspond to more inaccurate low-fidelities models. In all cases investigated below,  $\nu = 0.2$  and  $B = 2$  are used.

On the other hand, the high-fidelity model is constructed by adding an i.i.d. noise to the true value, i.e.,

$$z^H(x) = f(x) + \varepsilon(x), \quad (9)$$

where  $\varepsilon(x)$  is a white noise process. That is, its values for any  $x'$  values are i.i.d. Gaussian random variables with zero mean and variance  $\sigma_n^2$ , i.e.,  $\varepsilon(x') \sim \mathcal{N}(0, \sigma_n^2)$ . The variance  $\sigma_n^2$  is chosen to be 0.01, which is much smaller than the low-fidelity model discrepancy  $\delta$  in Eq. (8). One may also consider  $\varepsilon(x)$  as a Gaussian process that has a zero mean and a kernel with infinitely small correlation length scale  $l$  (see Eq. (2)). This assumption for the high-fidelity model is motivated by the fact that they are generally very accurate, with much smaller model discrepancies compared with their low-fidelity counterparts. For example, in the case of properly resolved LES with properly specified initial and boundary conditions, the simulation errors mainly come from finite time averaging, i.e., the procedure used to obtain integral or steady-state quantities from unsteady solutions. If the random uncertainty in the high-fidelity model becomes larger, the model discrepancy would no longer go to zero at the high-fidelity data points. Therefore, the assumptions adopted here neglect model bias, but the random model error mimics the statistical convergence of the unsteady simulations to a mean value. In practical CFD applications, under-resolved or otherwise improperly conducted LES may also lead to large errors, for which case a synthetic model similar to Eq. (8) can be used as well. This scenario is not pursued in this study.

### 3.1.3. Choice of Parameters

In all four test cases the uncertainty distributions for the input parameters are assumed to be Gaussian. In the one-dimensional cases (1 and 2), the input variable has a distribution  $x \sim \mathcal{N}(0.5, 0.25)$ . In the two-dimensional cases (3 and 4), the two input variables  $x_1$  and  $x_2$  are assumed to be independent, both having a normal distribution  $\mathcal{N}(0.5, 0.25)$ .

To obtain the ensemble  $\mathcal{D}^L$  for the low-fidelity model, the Latin hypercube sampling method is used to generate  $N^L = 500$  samples, which is confirmed to sufficiently represent the input uncertainty. The histogram of the samples is shown in Fig. 7. The number of samples required in ensemble  $\mathcal{D}^H$ , on which high-fidelity simulations are performed to obtain data for the construction of GPs for model discrepancies, varies depending on the dimension of input parameter space and the complexity of input-output the response surfaces. The choices for the four cases are presented in Table 1. Finally, the number  $N_c$  of realizations drawn from the posterior Gaussian process for the model discrepancy  $\delta$ , as shown in Eq. (5), is chosen to be 30 for all cases. This choice is found to give statistically converged results. The parameters

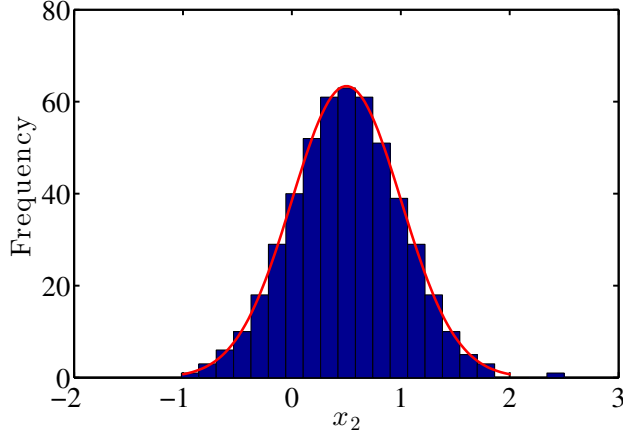


Figure 7: Histogram of the generated samples in the ensemble  $\mathcal{D}^L$  for low-fidelity model with comparison to the probability density function of the specified input uncertainty distribution  $x \sim \mathcal{N}(0.5, 0.25)$ .

that are common to all four cases presented below are summarized in Table 2.

case number	1	2	3	4
$N^H$	5, 2	5, 10, 20	10, 20	20, 40

Table 1: The number  $N^H$  of samples in the high-fidelity simulation ensemble  $\mathcal{D}^H$  for the four cases investigated. The number of samples required in the high-fidelity simulation depends on the dimension of the input parameter space and the complexity of the input–output the response surfaces. Various numbers are investigated for each case to demonstrate the effects of  $N^H$ .

### 3.2. Numerical Results

#### 3.2.1. Case 1: Monotonic 1D Response Surface

In this case, the monotonic 1D response surface as described in Eq. (6) with  $x_1 = 1.25$  is adopted as the truth, a comparison of which with the low-fidelity model results is shown in Fig. 8. The proposed multi-model method is performed to propagate the input uncertainty while also accounting for the model discrepancy. The response surface of model discrepancy is first constructed with a GP model. With the high-fidelity model evaluations, the assumed prior of the model discrepancy is updated with Bayesian inference to obtain the posterior GP. The prior and the posterior are presented in Fig. 2a and 2b, respectively.

parameter	$\nu$	$B$	$N^L$	$N_c$
value	0.2	2	500	30

Table 2: Common model parameters and the number of samples used in all cases.  $\nu$  and  $B$  are the proportionality constant and fixed constant, respectively, in the low-fidelity model discrepancy in Eq. (8);  $N^L$  is the number of samples in the low-fidelity simulation ensemble  $\mathcal{D}^L$ ;  $N_c$  is the number of realizations drawn from the posterior Gaussian process as shown in Eq. (5).

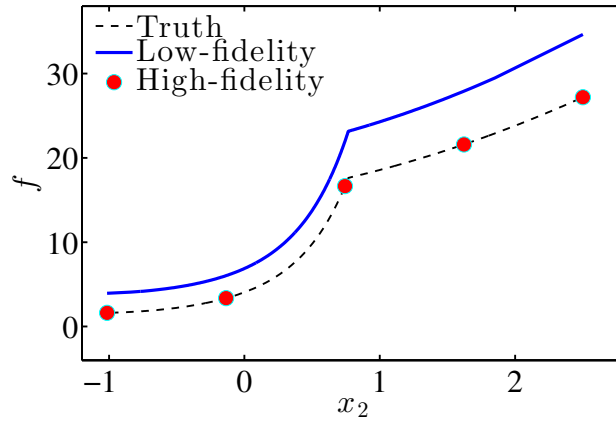


Figure 8: Response surface of monotonic mapping  $f_1(x_2)$  compared with low-fidelity model results  $z^L(x_2)$  with  $x_1$  being fixed as 1.25. Red dots represent high-fidelity model results.

Figure 2a shows an uninformative prior with relatively large confidence intervals due to the lack of information. Given the “observed” model discrepancy data calculated from the high-fidelity model results at the parameters in ensemble  $\mathcal{D}^H$ , the confidence interval of the posterior GP as shown in Fig. 2b is significantly reduced and the bias error has also been corrected. It is noteworthy that the confidence interval is small (but not zero) close to the observation points and increases further away from these points. In most part of the domain, the confidence interval covers the truth, but it fails between  $x_2 = -0.5$  and  $0.8$ , where the inferred distribution of  $\delta$  becomes slightly over-confident. This is due to the fact that the observation data are too sparse to capture the change of characteristics at around  $x_2 = 0.7$ . Moreover, the modular Bayesian approach, in which the hyperparameters are determined by the MLE method, only accounts for the most likely hyperparameter pair and ignores the other less likely possibilities. This inevitably leads to overconfidence, particularly in the cases where available data for inferences are sparse, and thus many possibilities are equally likely. However, for this trivial case the prediction error caused by the over-confidence is relatively minor compared to the absolute values of the model discrepancy functions, and thus its influences on the output uncertainty distributions are negligible. With this constructed model discrepancy GP, correction to the ensemble of low-fidelity model results can be conducted by  $N_c$  realizations from the GP.

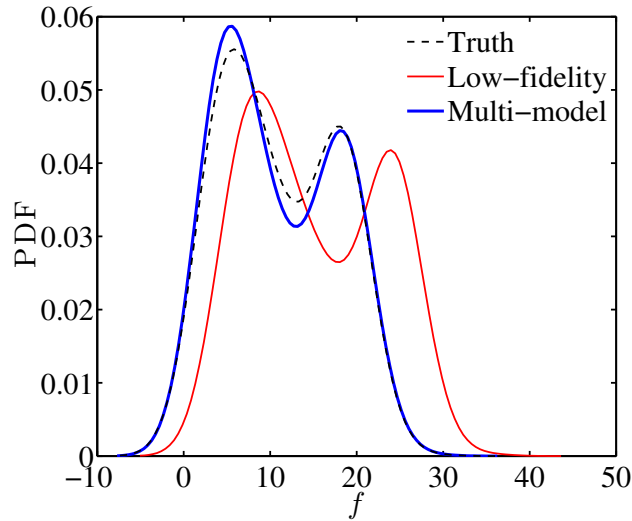


Figure 9: Corrected propagated uncertainty (blue bold solid line) based on the multi-model method compared with original results (red solid line) from the low-fidelity model.

The output ensembles after the correction are aggregated, and their Probability Density

Functions (PDF) are plotted in Fig. 9 with comparison to the truth and that obtained with the low-fidelity model alone. Rigorous, quantitative assessment and comparison of predicted distributions (in the form of PDFs or CDFs) are not straightforward and can be challenging. In this proof-of-concept study we only conduct approximate comparison via visual observations, which can be inadequate when the quality of two predicted distributions are not obvious. In those cases, more advanced metrics for comparing probability distributions are required, e.g., Area Validation Metric [38] and its modified variant [39]. Figure 9 shows that the normally distributed input uncertainty is mapped to a bi-modal distribution in the output uncertainty distribution. Comparing results in this figure, it is clear that the PDF of the corrected ensemble almost coincides with the corresponding truth, which indicates that the nonlinear model discrepancy has been reconstructed and corrected reasonably well with only five data points from high-fidelity simulations. A minor difference on the left peak and the valley of the PDF is due to the slight overconfidence of the inferred posterior GP as discussed above. This simple case demonstrates that the proposed multi-model scheme can effectively combine the results of models with multiple fidelities to improve uncertainty propagation process.

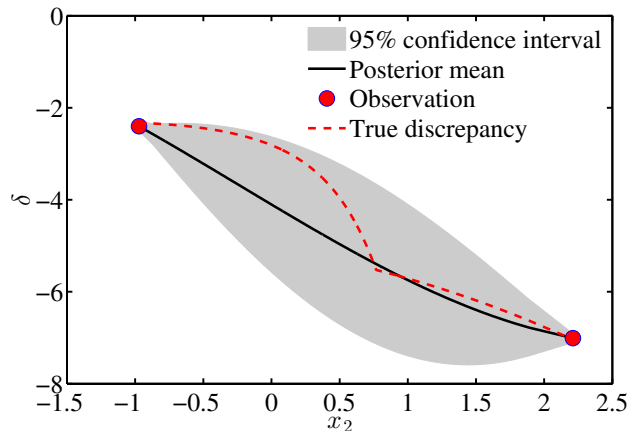


Figure 10: GP Modeling for model discrepancy with only two observations.

To investigate the influence of the amount of available data on the propagated uncertainty, we consider a relatively extreme case where the high-fidelity model is very expensive, and thus only two simulations can be afforded. Figure 10 shows that only two model discrepancy observations are obtained at the two ends of the input parameter interval. It can be seen that the confidence interval of posterior GP between the two observation points is much larger than that of the previous case with five observations. This increased uncertainty is due to



the smaller amount of available information (i.e., the observation data from the high-fidelity simulations).

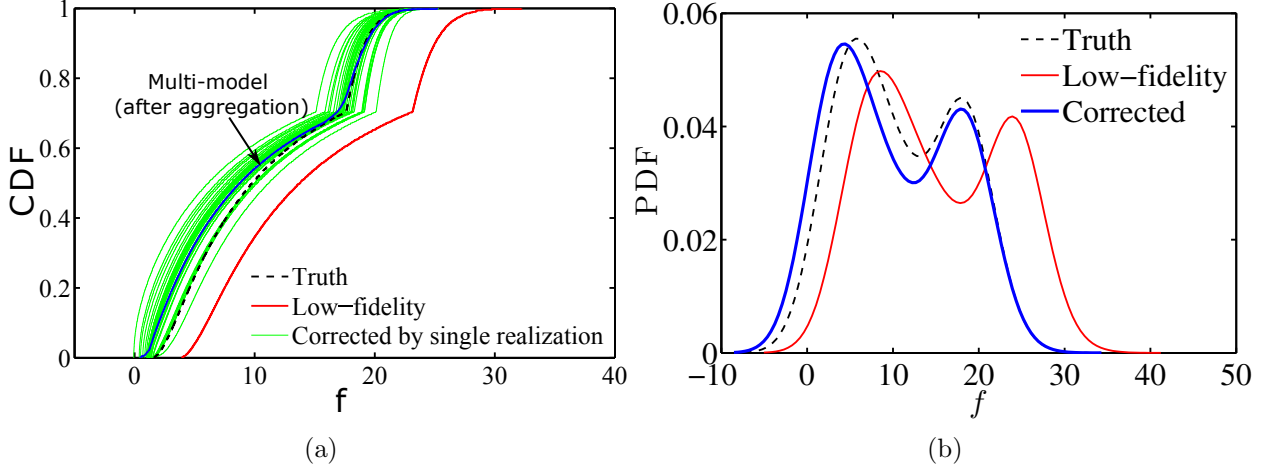


Figure 11: Propagated uncertainty based on the proposed framework compared with the results from the low-fidelity model. (a) Individual CDFs of QoI corrected obtained by using 30 realizations of the GP (before aggregation), as well as the aggregated CDF of the corrected QoI ensemble. (b) Aggregated PDF of the corrected QoI ensemble compared with the low-fidelity results. PDFs corresponding to the individual realizations (i.e., counterpart of the CDFs in panel (a)) are omitted to avoid cluttering the figure.

Figure 11 shows the distribution of the QoI ensemble corrected by using two observations. In Fig. 11a, a group of CDFs of QoI ensembles corrected by individual GP realizations before the aggregation are shown as green thin lines. It can be seen that the ensemble of corrected CDFs corresponding to different realizations in the posterior GP cover a probability region. The region covered by scattered CDF ensemble covers the truth, and corrects the model bias compared with that of the original low-fidelity model results. Note that we use a large number of individual realizations to conduct many corrections, instead of using the predicted mean of GP to perform one deterministic correction. Hence, an ensemble of corrected CDFs is obtained instead of a single CDF, which explicitly accounts for the uncertainties introduced by model discrepancy and the lack of high-fidelity simulation data. Figure 11b presents the PDF of the corrected QoI ensemble after aggregation. We can see that the PDF of corrected QoI agrees better with the truth than do the low-fidelity model results. Compared with Fig. 9, it is clear that the PDF corrected by two observations covers a larger region of output

values than that of the previous case with five observation data points. Meanwhile, the distribution is flattened slightly compared to the truth. This means the output uncertainty is enlarged, or stated differently, the entropy of information decreases [40]. This increase of uncertainty is referred to as *uncertainty inflation* in this work. For example, the transition from a Gaussian distribution to a uniform distribution over a similar range represents an uncertainty inflation, since the variance of the distribution increases in the transition and the information entropy decreases. The uncertainty inflation in Fig. 11b compared to the truth is introduced during the reconstruction of the model discrepancy GP, which is due to the limited number of high-fidelity simulations available for the GP reconstruction. In the proposed multi-model strategy, the uncertainties introduced in the reconstruction of discrepancies are obtained based on GP assumptions, and they almost always inflate the propagated input uncertainty. This inflation can be reduced by using more information obtained from high-fidelity simulations. However, it is important to note that the shape distortions of the CDF or PDF of the corrected results compared with the truth are not only due to the uncertainty inflation mentioned above. The errors caused by the overestimation or underestimation of model uncertainty as well as the errors in the mean discrepancy function also contribute to mis-characterization of the propagated uncertainty distribution. Since the stationary kernel function and a rather small number of observation points from the high-fidelity simulations are used, the change of characteristics of the discrepancy function over the parameter space are difficult to capture. Therefore, realizations from the posterior GP of model discrepancy are mostly smooth. In contrast, the two-regime feature that is clearly visible in the true discrepancy function. These errors are propagated to the output uncertainty distribution along with the quantified model uncertainty. Identifying and distinguishing the factors that distort the propagated uncertainty are far from trivial, and thus they will be further discussed in Section 4.2.

### 3.2.2. Case 2: Non-monotonic 1D Response Surface

A non-monotonic 1D mapping with more complex model discrepancy is considered. Its response surface compared with the low-fidelity model result is shown in Fig. 12, where the high-fidelity simulation data are plotted as well. It can be seen that the accuracy of the low-fidelity model results varies dramatically at different input locations, which poses a serious challenge for the reconstruction of model discrepancies.

Figure 13 shows the GP posterior of the model discrepancy. The prior GP is similar to that in Fig. 2a and it is thus omitted here. Not unexpectedly, most of model bias has been corrected based on only five observations compared with the initial guess of discrepancy, i.e.,

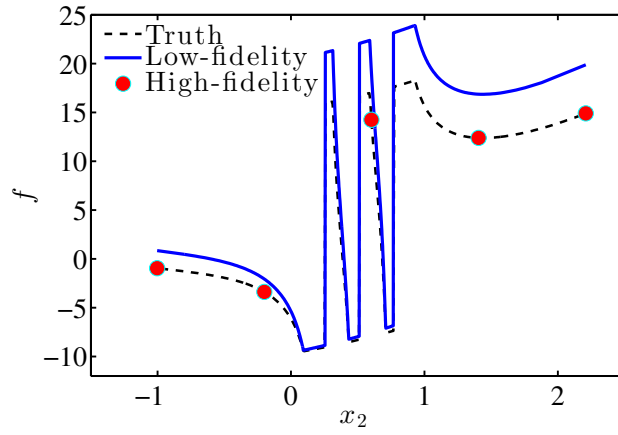


Figure 12: Response surface of the non-monotonic mapping  $f_2(x_2)$  compared with low-fidelity model results  $z^H(x_2)$  with  $x_1$  being fixed as 1.25.

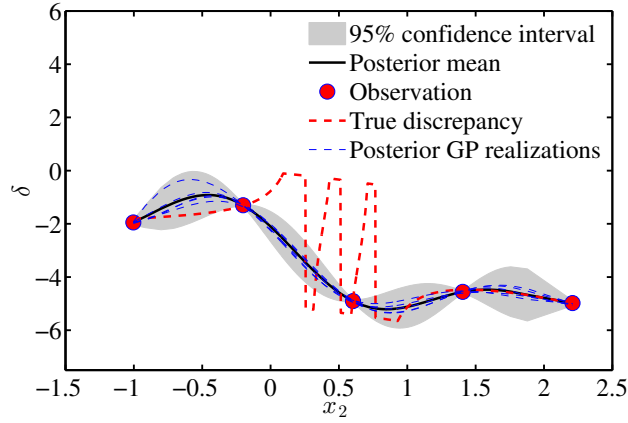


Figure 13: Posterior GP of the model discrepancy  $\delta$  obtained from Bayesian inference based on five observations from the high-fidelity model.

zero mean function. However, it is noted that the fluctuations between  $x = 0$  and 1 are not captured in the posterior mean. Furthermore, the confidence interval does not cover the truth in this region. The GP model constructed with the five data points is overconfident in this region. The reason is that it is very difficult to infer such a complex response surface of the model discrepancy with so little information, i.e., only five data points from high-fidelity model simulations. As pointed out in [41], the GP-based approach is a Bayesian framework to combine information available and to quantify the uncertainty, but it cannot create new information. As mentioned above, this is also due to the modular Bayesian approach, where an intrinsic feature of the MLE is the tendency to be overconfident [36]. Moreover, for this non-stationary problem, where the smoothness of the model discrepancy response surface is not homogenous on the entire input parameter domain, a GP with a stationary kernel function as adopted here is not suitable, since it tends to compromise between the smooth and the fluctuating regions.

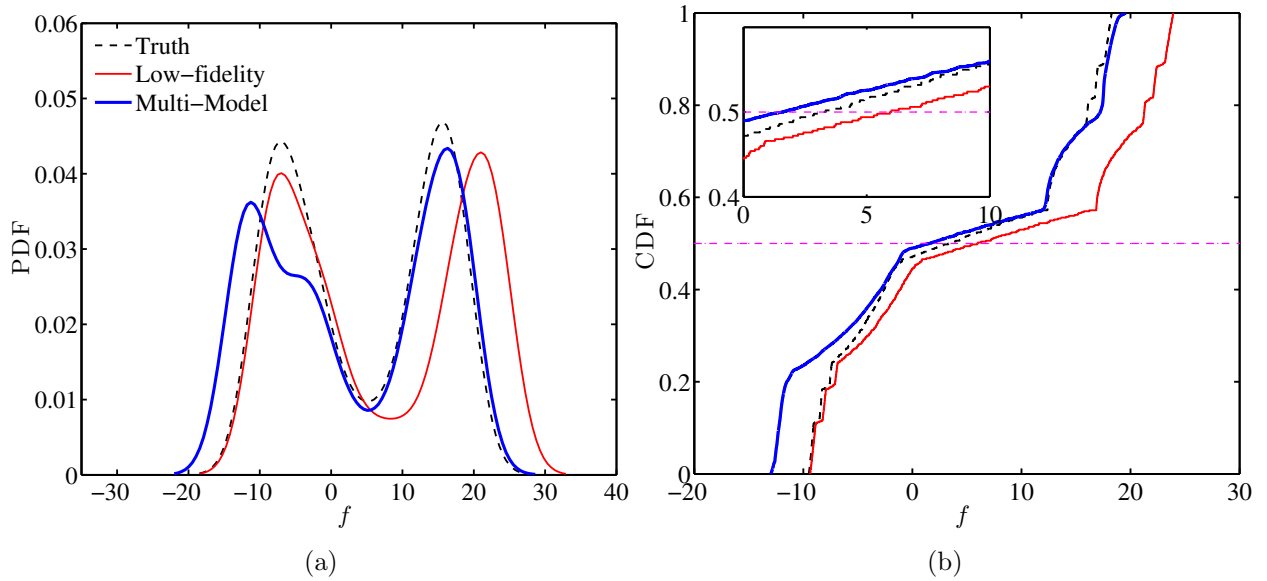


Figure 14: Propagated uncertainty based on the multi-model method compared with the results from the low-fidelity model. (a) Comparison of probability density functions. (b) Comparison of cumulative distribution functions.

Figure 14 presents the PDF and CDF of the corrected QoI uncertainty distribution after aggregation. Comparing the PDF of the low-fidelity model results with the truth in Fig. 14a, we can see that the shape of the PDF obtained with the low-fidelity model alone is also

different from the truth. In contrast, for the PDF obtained with the multi-model method, it is clear that the distribution at high QoI region ( $f_2 > 0$ ) is improved significantly, but the distribution at low QoI region ( $f_2 < 0$ ) actually deteriorated compared with the low-fidelity results. The unsatisfactory prediction of the QoI uncertainty distribution near the left peak is due to the overconfidence in the fluctuation region shown in Fig. 12. This can be improved by adopting a non-stationary GP model or adding more information, e.g., more data from high-fidelity simulations. Despite the unsatisfactory agreement with the truth in the lower QoI region, the bias in the mean value of  $f_2$  has been improved, which is shown in Fig. 14b. A horizontal dash-dotted line is used to indicate the cumulative probability equaling to 0.5. The value of  $f_2$  at which this line intersects with a CDF function is the median of this distribution. From Fig. 14b, it can be seen that the median of corrected QoI is improved compared with that of the low-fidelity model results, although the median of corrected QoI is over-corrected to some extent, which is due to the over-confidence of constructed model discrepancy. Moreover, it can be seen from the CDF that, beside the mis-characterization caused by over-confidence, uncertainty inflation of corrected results also occurs, which is due to the uncertainty in the construction of model discrepancy function.

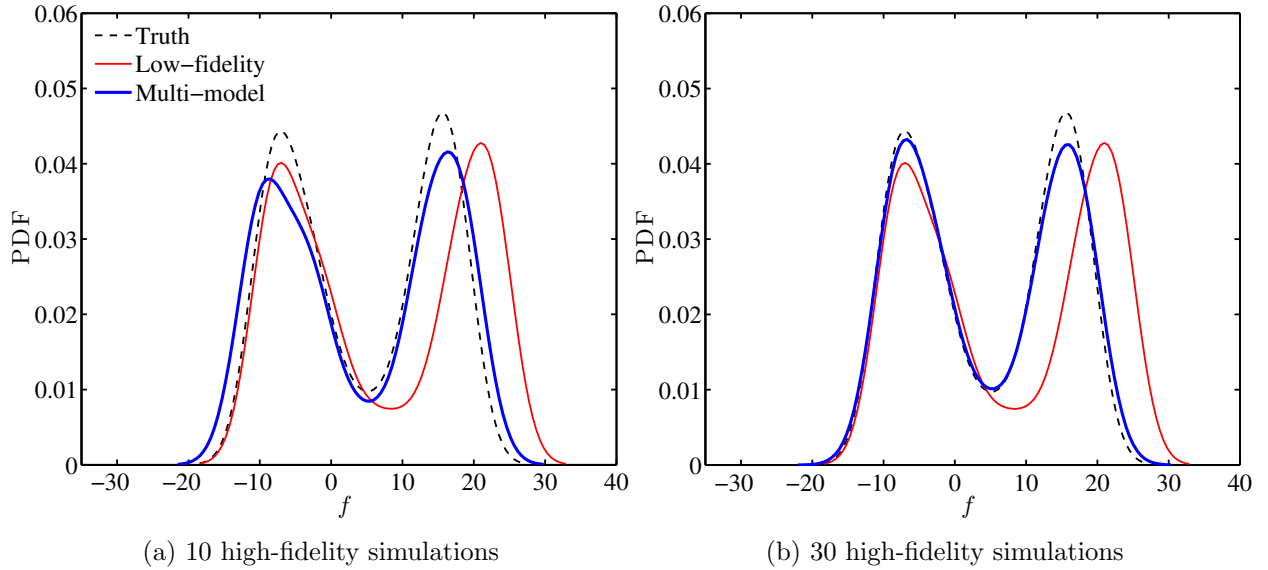


Figure 15: PDFs of corrected QoI ensembles by using the proposed method with (a) 10 high-fidelity simulations and (b) 30 high-fidelity simulations.

To further investigate the uncertainty inflation caused by data insufficiency, we studied

this case with a series of high-fidelity ensembles of different sizes. Because of space limitations, only the results with 10 and 30 high-fidelity simulations are presented in Fig. 15, and the other results that are omitted here had the same trend. Comparing Figs. 15a and 15b with Fig. 14a, it can be seen that increasing the number of high-fidelity simulations improves the predicted output uncertainty PDF, which is demonstrated by the decreased discrepancy with the truth. The propagated uncertainty after correction is inflated (the PDF curve is slightly flatter and wider), but all of the inflated uncertain regions can cover that of the truth, and this uncertain inflation can be reduced by adding more high-fidelity model results. It demonstrates that the GP-based Bayesian inference method does not create information, but it can effectively combine the information available; if more information comes in, the results can be improved.

### 3.2.3. Case 3: Monotonic 2D Response Surface

Uncertainty propagation with multiple input variables is much more difficult than problems with a 1D input parameter space. Because it takes many more samples to fill up a higher-dimensional space, the fact that is commonly referred to as “curse of dimensionality”, multiple dimensional problems need more information for Bayesian inference. This is a challenge for the proposed multi-model method. Figure 16 shows model discrepancy re-

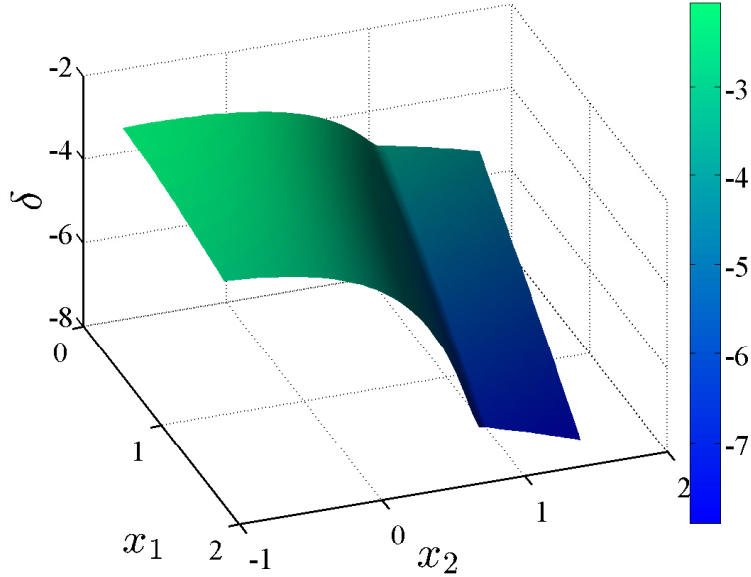


Figure 16: Model discrepancy for the low-fidelity model (see Eq. (8)) for the Case 3, the response surface of which is shown in Eq. (5).

sponse surface of the low-fidelity model for the monotonic 2D case. Compared to the model discrepancy in the corresponding 1D response surface of Case 1, the response surface and the model discrepancy in this case exhibit monotonic dependence on  $x_1$  in addition to the nonlinear dependence on  $x_2$ .

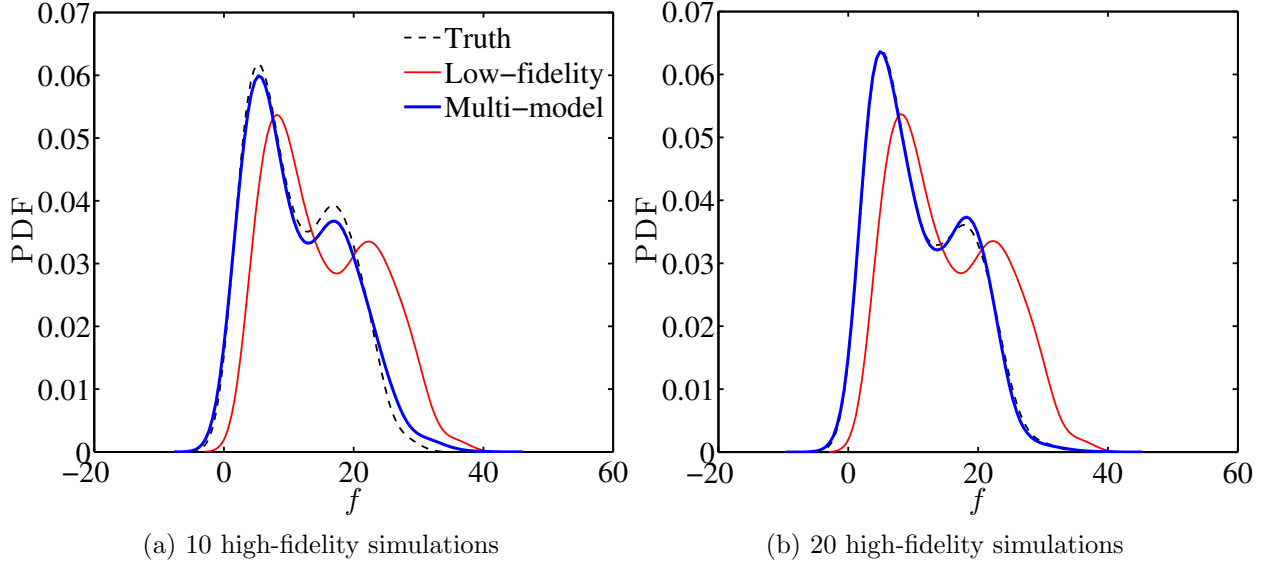


Figure 17: PDFs of the corrected QoI  $f_1(x_1, x_2)$  ensemble from the current multi-model method with (a) 10 high-fidelity simulations and (b) 20 high-fidelity simulations.

After the propagation process with the proposed multi-model method, PDFs of the propagated QoI ensembles are presented in Figs. 17a and 17b, showing multi-model results with 10 and 20 high-fidelity simulations, respectively. From Fig. 17a, it can be seen that the PDF of the low-fidelity model output ensemble is distorted by model discrepancy compared with that of the truth. With the correction obtained from 10 high-fidelity model results, the PDF of the corrected ensemble agrees well with the truth. We can see that the PDF of the corrected ensemble covers a slightly larger uncertainty region than the truth does, in line with previous findings in Cases 1 and 2. This uncertainty inflation is caused by insufficient information from the high-fidelity simulations. After increasing the high-fidelity simulations to 20, the uncertainty inflation is reduced and the new result agrees with the truth better. This can be clearly seen in Fig. 17b, where multi-model results are almost identical to the truth.

### 3.2.4. Case 4: Non-monotonic 2D Response Surface

Since the model discrepancy of the low fidelity model as assumed in Eq. (8) has a term  $\nu f$ , which is linearly proportional to the output quantity  $f$ , the response surface of the model discrepancy for this non-monotonic 2D case is also very complex with strong fluctuations and nonlinearity, which can be seen in Fig. 18. We applied the same multi-model approach, and the propagated output uncertainties are presented in Fig. 19, where the results obtained by using 20 and 40 high-fidelity simulations are shown.

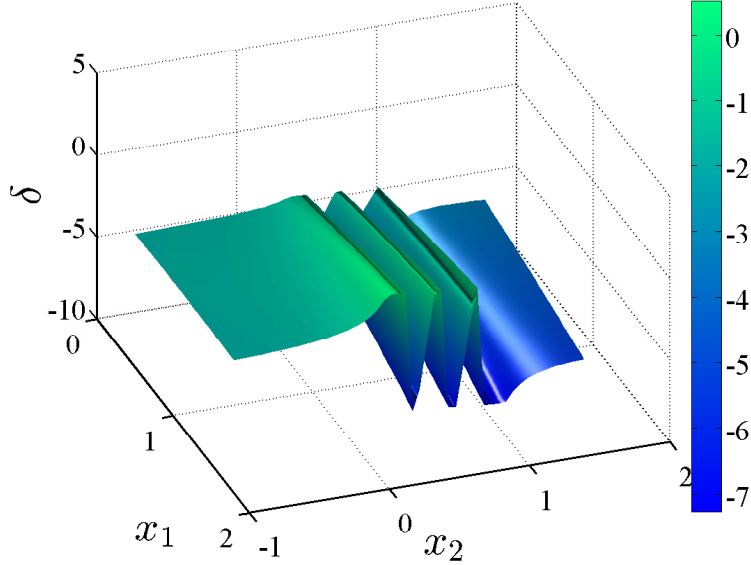


Figure 18: Response surface of model discrepancy for low-fidelity model.

It can be seen from Fig. 19a that with 20 high-fidelity model results, the PDF of the QoI ensemble is corrected to be closer to the truth, and especially the mean value agrees well with the truth. However, the uncertainty of corrected ensemble is larger than that of the truth, as is evident from the flatter and wider PDF of the corrected results, indicating uncertainty inflation. Compared with the results of the monotonic case, the uncertainty inflation and mis-characterization here are more significant. This is because the response surface for model discrepancy is much more complex than that of the monotonic case. Consequently, more information is needed to reconstruct this complicated surface. The lack of information is reflected as model uncertainty, which is quantified by the GP model and subsequently inflates the propagated uncertainty in the QoI. Meanwhile, the lack of high-fidelity model data also causes the reconstruction procedure to miss some important feature of response surface, which then leads to mis-characterizing the propagated uncertainty. This can be



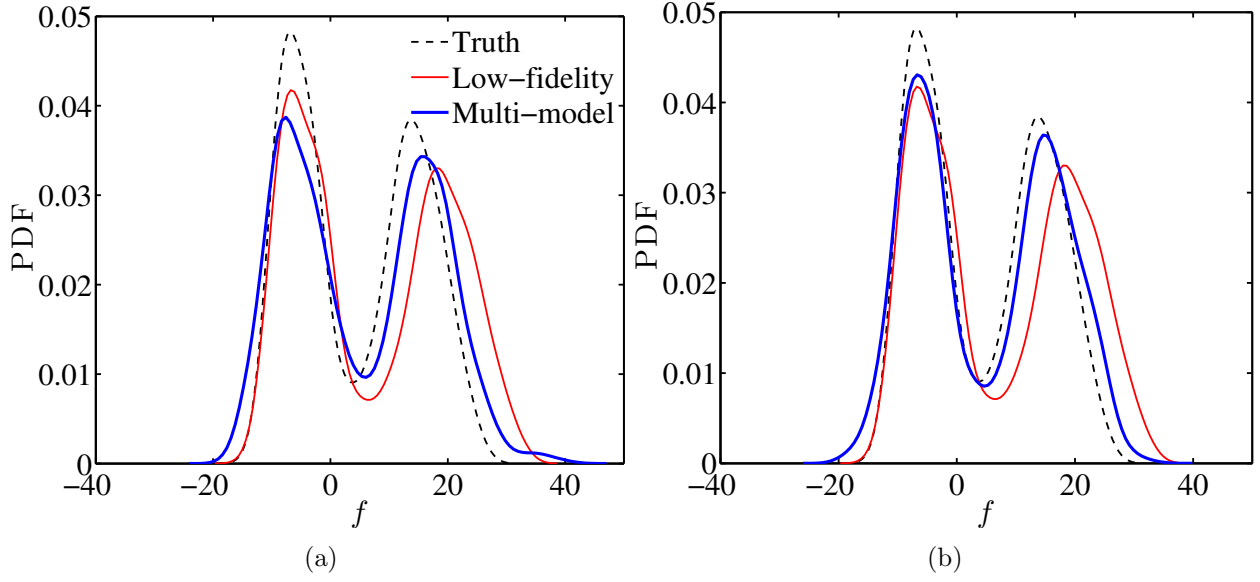


Figure 19: PDFs of corrected QoI  $f_2(x_1, x_2)$  ensemble from current multi-model method with (a) 20 high-fidelity samples and (b) 40 high-fidelity samples.

illustrated by comparing Figs. 19a and 19b, which show the PDFs of corrected ensembles with 20 and 40 observations, respectively, of the model discrepancy. It is clear that with more high-fidelity simulations the agreement between the predicted QoI uncertainty and the truth improves, and the corresponding uncertainty inflation decreases. However, the improvements are not significant. In this non-monotonic 2D case, even with 40 high-fidelity simulations the distribution of the corrected QoI ensemble still is not identical to the truth, and the uncertainty inflation cannot be completely eliminated as in the monotonic case. This can be attributed to the fact that the GP model with a stationary kernel function cannot adequately handle problems with very different characteristics in different regions of the input parameter space. Specifically, with a stationary kernel, identification of whose hyperparameters is based on MLE, the GP construction tends to compromise between regions of different characteristics [32], leading to mis-characterizations of propagated QoI uncertainty. Although the inflation can be reduced by increasing the number of high-fidelity simulations, the mis-characterization becomes a dominant factor that deteriorates the propagated QoI uncertainty. A non-stationary kernel with a careful design of observation locations can be used to improve the results in these situations.

## 4. Discussion

### 4.1. Advantage Over Single-Model Approaches

When evaluating the merits of the proposed multi-model approach, there are two legitimate questions to ask before the proposed method can be justified. (1) What benefits does the high-fidelity model offer? Or in other words, would it be better to simply allocate all computational resources to low-fidelity simulations? (2) What benefits does the low-fidelity model offer? Or in other words, would it be better to allocate all computational sources to high-fidelity simulations? Inevitably, answers to these questions are not straightforward, and ultimately they depend on the relative merits and costs of the respective models. Consequently, investigations of these issues can only be based on specific low-fidelity and high-fidelity models. In the following we will discuss the two issues above based on the models assumed in Eqs. (8) and (9). The first question above on the benefits of the high-fidelity model has been answered clearly in Section 3.2. When the low-fidelity model alone is used in the uncertainty propagation, the obtained output uncertainties (PDFs and CDFs) show significant bias, which are due to the prediction bias of the low-fidelity models. In contrast, when data from high-fidelity simulations are incorporated into the uncertainty propagation process, even though the amount of available data is very limited (e.g., only two data points in Case 1 (Fig. 10)), the improvements on the output uncertainty are clearly visible. This observation effectively demonstrates the value of the high-fidelity model, without which the correct output uncertainty cannot be obtained, no matter how many samples are used in the low-fidelity model evaluations. This is illustrated in Fig. 20, where the PDF obtained by using the low-fidelity model alone shows a significant bias even with a very large ensemble with 500 samples. The second question is closely related to the investigations conducted by Higdon et al. [23] regarding the usefulness of numerical simulations in a prediction method combining numerical simulations and experimental data. They concluded that the numerical simulations have significant contributions in reducing the prediction uncertainties, and thus the combined simulation–data based framework is superior to the pure experimental data-driven approach. In this work we carry out a similar investigation in the context of uncertainty propagation by comparing the multi-model results with those obtained by utilizing high-fidelity simulations alone. That is, in this alternative approach, data obtained from high-fidelity simulations are used to fit a GP, from which samples are drawn to propagate the input uncertainties. The low-fidelity-mode is not used in this alternative approach.

Results from Case 1, i.e., the simplest monotonic 1D response surface case discussed above in Sec. 3.2.1, are presented below to illustrate the benefit of the low-fidelity model.

For this comparison, five high-fidelity model results are adopted for both methods and the same parameters as above are used. The PDF of the QoI uncertainty distribution obtained by using the multi-model approach and that obtained by using the high-fidelity model alone are shown in Fig. 20. From this figure, one can see that the multi-model results agree with the truth extremely well. In contrast, the results from the high-fidelity model alone deviate significantly from the truth. Although the high-fidelity model based approach indeed corrected the obvious bias that is present in the pure low-fidelity model results, it fails to capture the bi-modal feature of output uncertainty distribution, which is an essential feature present in the true QoI distribution. Overall, the multi-model method gives much better results for the QoI uncertainty than the single-model approach based on the high-fidelity model alone.

In summary, the comparison above suggests that, by combining the low- and high-fidelity models, the proposed approach produces better results than either model can yield alone.

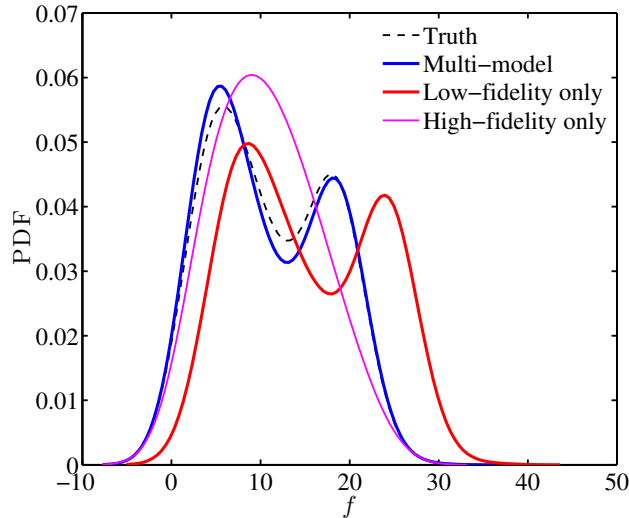


Figure 20: Propagated uncertainty distribution obtained by using the proposed multi-model method compared with those obtained by using the single-model approaches, i.e., with either high-fidelity model or the low-fidelity model alone.

#### 4.2. Analysis of Uncertainty Distortion in the Multi-model Strategy

From the results of the four test cases presented in Section 3, we can see that the propagated QoI uncertainty distributions are often distorted in the low-fidelity model results, which is due to the discrepancy of the model used in the uncertainty propagation. As mentioned in

Section 1.3, the coupling between the input uncertainty and the model-form uncertainty is very complex, which jointly determines the obtained QoI uncertainty distribution. When the proposed multi-model strategy is used for uncertainty quantification, some of the model-form uncertainties are removed, but interpolation errors due to the limited number of high-fidelity simulations are introduced in, which further complicate the situation. It is thus helpful to discuss the distortions of the propagated QoI uncertainty both in the single-model approach and in the multi-model approach.

To facilitate discussion, a simple linear response surface is used as example. However, despite its simplicity, the linear response surface example is representative in that it can be considered as an approximation of a generic, nonlinear response surface in a small (or infinitesimal) region of the input parameter space.

Although the interaction between the model-form uncertainty and the input uncertainty can lead to complex distortions on the QoI uncertainty distribution, we propose to decompose the distortions caused by the model discrepancy into three basic forms: bias, inflation, and deflation, which are illustrated in Figs. 21a–c, respectively. In Fig. 21a, the model used for uncertainty propagation has a constant bias compared with the true response surface, which leads to a bias of the same amount in the PDF of the QoI uncertainty. In Fig. 21b the model response surface has a steeper slope than that of the truth, which leads to a flatter and wider PDF compared with that of the true QoI uncertainty. This is referred to as “inflation”. Figure 21c shows an opposite scenario, where the modeled response surface has a milder slope than that of the truth, leading to a QoI uncertainty distribution with “deflation”, indicated by the PDF that is more concentrated than the true QoI uncertainty PDF. Of course, in general the model response surface can have different slopes and biases in different regions, leading to a QoI uncertainty distribution with a combination of all three forms of distortion compared with the truth. This generic case is illustrated in Fig. 21d.

The proposed multi-model method aims to remove the above-mentioned distortions caused by the discrepancy of the low-fidelity model through the use of high-fidelity simulation data, or more precisely, through a GP of the model discrepancy constructed from the data. Conceptually, the proposed method can be considered as consisting of two steps as follows, although the distinction between the two steps is not made in actual implementation:

1. **Model calibration:** First, the low-fidelity model results are corrected with the posterior mean of the model discrepancy GP. The correction in this step is purely deterministic, and can be considered as implicit way of calibrating or correcting the low-fidelity model simulation results with the high-fidelity simulation data. Ideally, if the amount

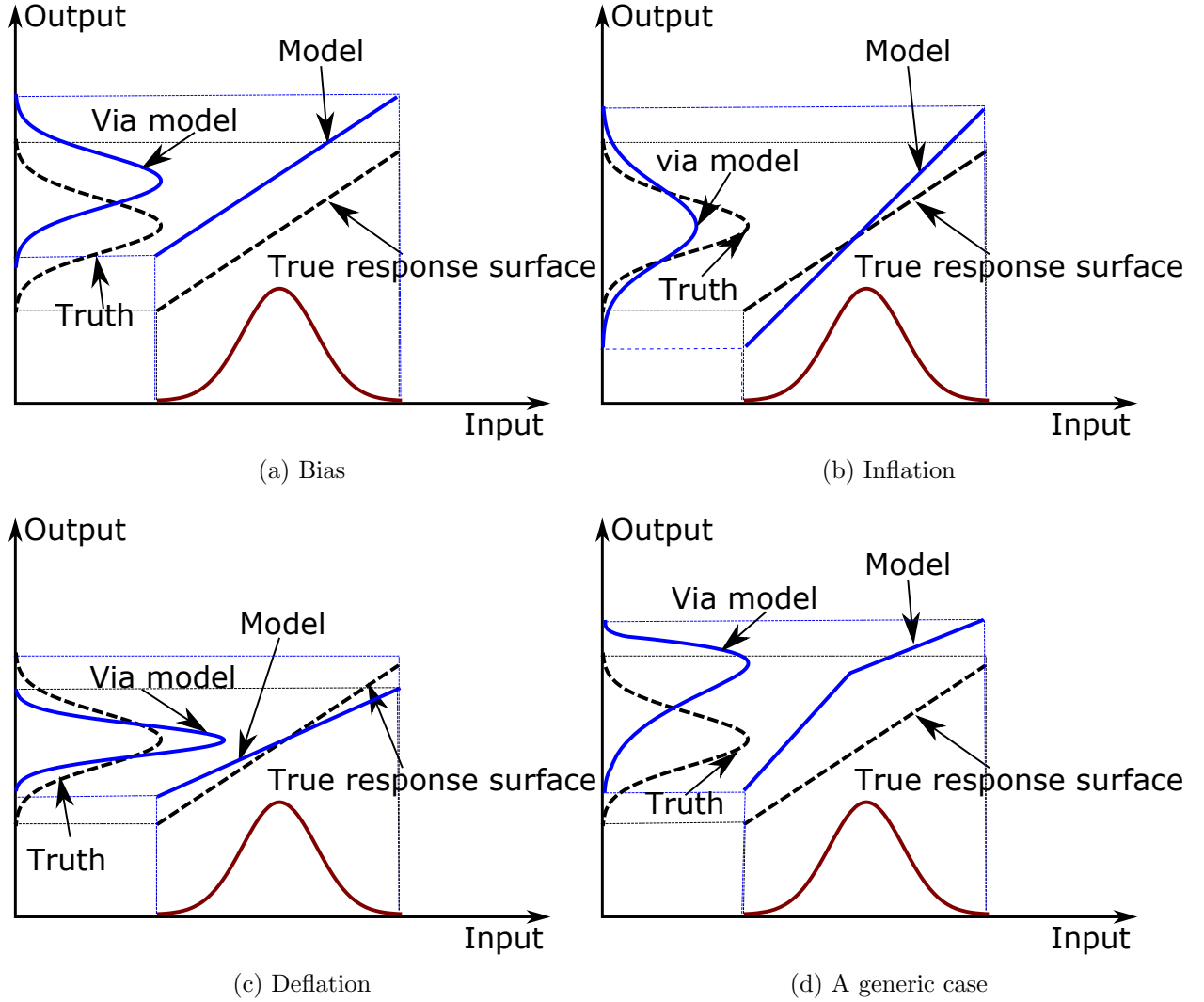


Figure 21: Three basic forms of distortions on the QoI uncertainty distribution caused by the discrepancy of the model used in the uncertainty propagation : (a) bias, (b) inflation and (c) deflation. Uncertainty distortion in a generic case often has a combination of these three basic forms, which is shown in panel (d).

of high-fidelity simulation data were sufficient and the constructed GP could accurately represent the true model discrepancy, this step would eliminate the distortions caused by the low-fidelity model. In practice, we indeed expect and have observed in Section 3.2 that most of the distortions caused by the low-fidelity models were removed as illustrated in Fig. 22 (see the QoI uncertainty PDF annotated with “mean

correction”). Inevitably, however, in most cases high-fidelity model data are limited due to the associated high computational costs, and consequently the posterior mean of the discrepancy GP cannot capture all the small-scale features of the true response surface. As a result, the PDF with posterior mean correction still has distortions (bias, inflation, deflation, or a combination thereof), which is illustrated by the difference between the true PDF and the PDF with mean correction. In some cases, e.g., when the high-fidelity simulation data are very sparse and grossly misrepresent some features in the discrepancy function, it can even introduce more distortion than what is present in the low-fidelity results. Therefore, it is important to explicitly quantify the amount of uncertainty introduced by the sparseness of the high-fidelity simulation data, i.e., the interpolation uncertainties.

2. **Uncertainty quantification:** In the second step, the interpolation uncertainties are explicitly quantified based on the assumptions in GP. Hence, it can be seen that the confidence intervals as shown in Figs. 2, 11, and 13 are merely a reflection of the data sparseness, and are not a true representation of the model-form uncertainty. This uncertainty almost always flattens the QoI distribution PDF obtained with mean correction, leading to uncertainty inflation. One can see this clearly by comparing the PDF after “mean correction” and that with “aggregated correction” in Fig. 22. This is because a number of realizations from the GP accounting for various possibilities of the discrepancy are used to correct the low-fidelity results, which is in contrast to the PDF with mean correction, where only one possible discrepancy function (the posterior mean) is used for correction. In the cases where the posterior mean of the discrepancy GP misses some small-scale, minor features of the true discrepancy, the inflation introduced by the aggregation tends to reduce the over-confidence in the PDF with only mean correction, and thus this is indeed a faithful inflation of the QoI uncertainty.

Ideally, the constructed discrepancy GP only misses small-scale, insignificant features of the truth and is indeed able to capture all large-scale, important features. However, if the response surface of model discrepancy is very complex, possibly with regime-dependent characteristics and strong fluctuations (see Fig. 6), the constructed GP can miss critical features of the truth. In such cases some distortions caused by the low-fidelity model, which can be bias, inflation, deflation, or a combination thereof, would remain even after the GP-based corrections. We can refer to these inaccuracies as *residual distortions* due to the errors in the construction of the model discrepancy GP. Even though the interpolation uncertainties

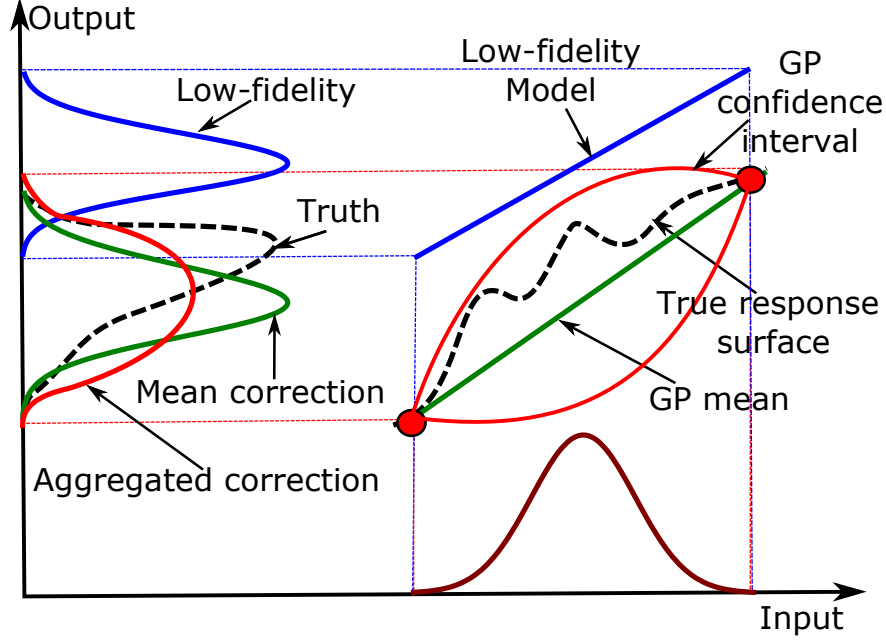


Figure 22: Schematic illustration of a two-step conceptual model of the proposed multi-model strategy for model calibration and uncertainty quantification. (1) Calibration: the QoI uncertainty distribution obtained with the low-fidelity model is corrected with the posterior mean of discrepancy GP, via which most of the distortions caused by low-fidelity model are removed. (2) Uncertainty quantification: the interpolation uncertainties associated with the sparseness of the high-fidelity simulation data are explicitly accounted for based on GP assumptions, which causes the PDF with “mean correction” to flatten and widen, leading to the PDF with “aggregated correction”.

caused by the sparseness of the high-fidelity simulation data tend to inflate the QoI uncertainty (i.e., from the PDF with “mean correction” to that with “aggregated correction”, see Fig. 22), the final QoI uncertainty distribution is not necessarily inflated compared with the truth, as the residual distortion can be a strong deflation and can dominate the final QoI uncertainty.

Figure 23 shows an extreme situation where the proposed multi-model strategy with a stationary GP model may fail. It can be seen that in the small region where the input uncertainty PDF concentrates, the true response surface has a sharp, abnormal bump. If there is no observation data in this region, which is very likely due to the small area the bump covers, the constructed GP will miss this important feature and become over-confident. Consequently, the propagated QoI uncertainty with the over-confident GP (green line in Figure 23) will be severely distorted compared to the truth (dashed line). Several strategies can be used to improve the results in this situation. First, increasing the amount of high-fidelity simulation data makes it more likely to successfully capture the complex features of the response surface. When the amount of data is constrained by practical considerations (e.g., computational expenses), experimental design methods [37] can be helpful in determining the locations of data points. Moreover, using physical insights or engineering intuition can also help improving the allocation of observation data in the parameter space. For example, more efficient use of high-fidelity simulation data can be achieved by placing more observation points in regions of the parameter space where a regime change is likely to occur (e.g., due to flow separation or laminar-to-turbulent flow transition) or small-scale features may be present (e.g., drastic change of lift coefficients due to stall; see Fig. 1). Finally, using a non-stationary GP that incorporates prior physical knowledge can be useful in avoiding the problem highlighted in Fig. 23, since stationary kernels tend to be over-confident for non-stationary response surfaces.

## 5. Conclusions

Model uncertainties play an important role in computational fluid dynamics and particularly in turbulent flow applications, where first-principle-based direct simulations are prohibitively expensive and thus are not affordable in most practical applications. In CFD applications, when propagating input uncertainties to output uncertainty distributions, uncertainties due to the numerical model used in the uncertainty propagation can distort the obtained output uncertainty. The effects of the model uncertainties on the results should be properly identified, while, at the same time, sufficient samples should be used to avoid



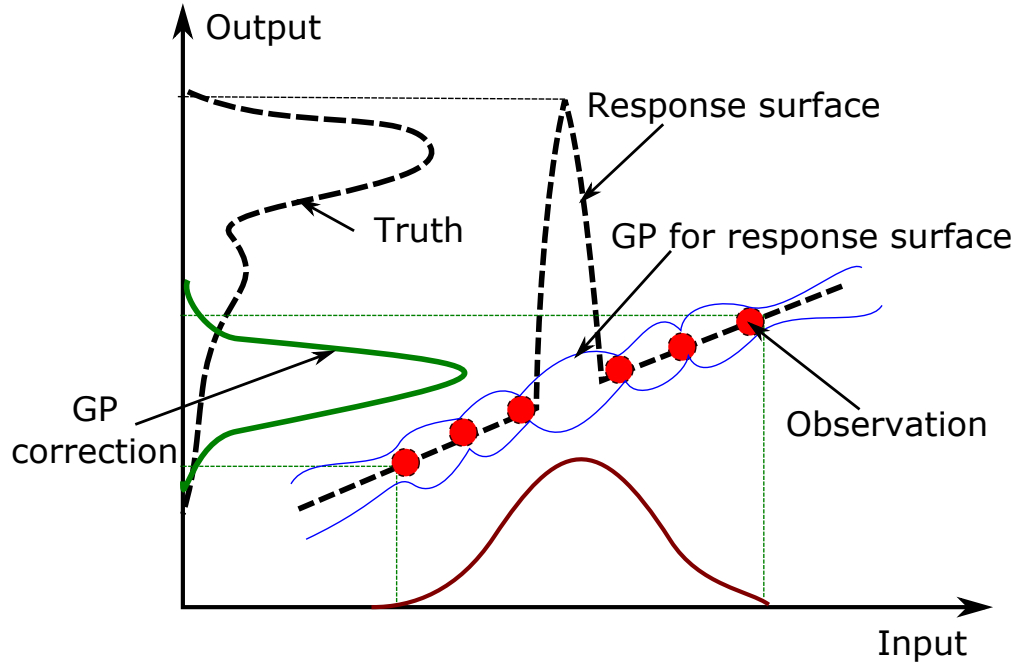


Figure 23: An extreme situation where correction with model discrepancy constructed with a stationary GP can cause severe distortion in the QoI uncertainty. A sharp, abnormal bump in the response surface is present in the region of the input parameter space where input uncertainty has high probability. The observations are too sparse to capture this feature, leading to over-confidence in the constructed GP and distortion in the QoI uncertainty.

sampling errors. In this work we propose a multi-model uncertainty propagation strategy, where model uncertainties are accounted for by using numerical models of different fidelities. Specifically, a computationally cheap, low-fidelity model is used to propagate a large ensemble of the input uncertainty distribution, while a computationally expensive, high-fidelity model is used to propagate a small number samples. The data obtained from the high-fidelity simulations are used to construct a probabilistic response surface of the model discrepancy. The construction is performed using a Bayesian framework and by assuming that the model discrepancy can be described by a Gaussian process. Samples from the constructed probabilistic response surface are used to correct the output ensemble propagated by the low-fidelity model and thus correcting the associated biases and distortions. This response surface is the posterior distribution of the model discrepancy and is also a Gaussian process. Compared to previous work, the proposed method is based on a smaller number of assumptions. In particular, Gaussian processes are not assumed for the low- and high-fidelity models themselves and only for the model discrepancy. We argue that this assumption is more realistic for practical CFD applications. Moreover, while previous authors used the GP based multi-model framework mostly for predictions, in this work, the multi-model framework is used for uncertainty propagation problems of relevance to CFD applications.

Four synthetic test cases with assumed input–output response surfaces and low- and high-fidelity models that are qualitatively similar to those in CFD applications are used to demonstrate the merits of the proposed method. Simulation results indicate that biases and distortions of the propagated output uncertainty, which are caused by the inaccuracy of the low-fidelity model or by the low number of samples available from high-fidelity model, can be significantly reduced by incorporating data from the high-fidelity simulations through the multi-model framework. Inflation in the propagated output uncertainty associated with the model uncertainty is quantified. By incorporating more data from the high-fidelity simulations, the uncertainty inflation due to the model uncertainty can be reduced. Overall, the results clearly demonstrate that, by combining the low- and high-fidelity models, the proposed multi-model uncertainty propagation strategy leads to significantly improved results compared with what either model can achieve individually. The proposed multi-model strategy can be extended to problems of multiple spatial and/or parametric dimensions, and thus it is a promising method for uncertainty propagation in CFD applications.

## References

- [1] O. P. Le Maitre, O. M. Knio, *Spectral Methods for Uncertainty Quantification: with Applications to Computational Fluid Dynamics*, Springer, 2010.
- [2] W. L. Oberkampf, T. G. Trucano, C. Hirsch, Verification, validation, and predictive capability in computational engineering and physics, *Applied Mechanics Reviews* 57 (5) (2004) 345–383.
- [3] C. J. Roy, W. L. Oberkampf, A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing, *Computer Methods in Applied Mechanics and Engineering* 200 (25-28) (2011) 2131–2144.
- [4] P. H. Garthwaite, J. B. Kadane, A. O’Hagan, Statistical methods for eliciting probability distributions, *Journal of the American Statistical Association* 100 (470) (2005) 680–701.
- [5] M. Beer, S. Ferson, V. Kreinovich, Imprecise probabilities in engineering analyses, *Mechanical Systems and Signal Processing* 37 (1) (2013) 4–29.
- [6] S. Asmussen, P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer, 2010.
- [7] J. C. Helton, F. J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, *Reliability Engineering & System Safety* 81 (1) (2003) 23–69.
- [8] R. G. Ghanem, P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, revised Edition, Dover Publications, 2003.
- [9] H. N. Najm, Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics, *Annual Review of Fluid Mechanics* 41 (2009) 35–52.
- [10] J. A. Hoeting, D. Madigan, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: a tutorial, *Statistical Science* 14 (4) (1999) 382–401.
- [11] A. E. Raftery, M. Kárný, P. Ettler, Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill, *Technometrics* 52 (1) (2010) 52–66.

- [12] A. Barth, C. Schwab, N. Zollinger, Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients, *Numerische Mathematik* 119 (1) (2011) 123–161.
- [13] S. Mishra, C. Schwab, Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data, *Mathematics of Computation* 81 (280) (2012) 1979–2018.
- [14] M. B. Giles, Multilevel Monte Carlo path simulation, *Operations Research* 56 (3) (2008) 607–617.
- [15] S. Mishra, C. Schwab, J. Šukys, Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions, *Journal of Computational Physics* 231 (8) (2012) 3365–3388.
- [16] F. Müller, D. W. Meyer, P. Jenny, Solver-based vs. grid-based multilevel monte carlo for two phase flow and transport in random heterogeneous porous media, *Journal of Computational Physics* 268 (2014) 39–50.
- [17] A. Narayan, Y. Marzouk, D. Xiu, Sequential data assimilation with multiple models, *Journal of Computational Physics* 231 (1) (2012) 6401–6418.
- [18] A. Narayan, C. Gittelsohn, D. Xiu, A stochastic collocation algorithm with multifidelity models, *SIAM Journal on Scientific Computing* 36 (2) (2014) A495–A521.
- [19] X. Zhu, A. Narayan, D. Xiu, Computational aspects of stochastic collocation with multifidelity models, *SIAM/ASA Journal on Uncertainty Quantification* 2 (1) (2014) 444–463.
- [20] N. Cressie, *Statistics for Spatial Data*, Wiley-Interscience New York, 1993.
- [21] A. I. Forrester, A. Sóbester, A. J. Keane, Multi-fidelity optimization via surrogate modelling, *Proceedings of the royal society A: mathematical, physical and engineering science* 463 (2088) (2007) 3251–3269.
- [22] M. C. Kennedy, A. O’Hagan, Predicting the output from a complex computer code when fast approximations are available, *Biometrika* 87 (1) (2000) 1–13.

- [23] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafo, R. D. Ryne, Combining field data and computer simulations for calibration and prediction, *SIAM Journal on Scientific Computing* 26 (2) (2004) 448–466.
- [24] L. Le Gratiet, Bayesian analysis of hierarchical multifidelity codes, *SIAM/ASA Journal on Uncertainty Quantification* 1 (1) (2013) 244–269.
- [25] I. Sahin, J. W. Crane, K. P. Watson, Application of a panel method to hydrodynamics of underwater vehicles, *Ocean engineering* 24 (6) (1997) 501–512.
- [26] D. C. Wilcox, *Turbulence Modeling for CFD*, 2nd Edition, DCW Industries, 1998.
- [27] P. Sagaut, *Large Eddy Simulations for Incompressible Flows: An Introduction*, Springer, 2006.
- [28] I. H. Abbott, *Theory of wing sections, including a summary of airfoil data*, Courier Dover Publications, 1959.
- [29] W. McCroskey, K. W. McAlister, L. Carr, S. Pucci, An experimental study of dynamic stall on advanced airfoil sections. volume 1. summary of the experiment., Tech. rep., DTIC Document (1982).
- [30] D. Huang, T. Allen, W. Notz, R. Miller, Sequential kriging optimization using multiple-fidelity evaluations, *Structural and Multidisciplinary Optimization* 32 (5) (2006) 369–382.
- [31] C. K. Williams, C. E. Rasmussen, *Gaussian processes for machine learning*, MIT Press, 2006.
- [32] R. B. Gramacy, H. K. Lee, Bayesian treed gaussian process models with an application to computer modeling, *Journal of the American Statistical Association* 103 (483) (2008) 1119–1130.
- [33] Y. Xiong, W. Chen, D. Apley, X. Ding, A non-stationary covariance-based kriging method for metamodeling in engineering design, *International Journal for Numerical Methods in Engineering* 71 (6) (2007) 733–756.
- [34] S. Ba, V. R. Joseph, et al., Composite gaussian process models for emulating expensive functions, *The Annals of Applied Statistics* 6 (4) (2012) 1838–1860.

- [35] P. D. Arendt, D. W. Apley, W. Chen, Quantification of model uncertainty: Calibration, model discrepancy, and identifiability, *Journal of Mechanical Design* 134 (10) (2012) 100908.
- [36] O. Schabenberger, C. A. Gotway, *Statistical methods for spatial data analysis*, CRC Press, 2004.
- [37] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, *Statistical science* (1989) 409–423.
- [38] W. L. Oberkampf, C. J. Roy, *Verification and Validation In Scientific Computing*, Cambridge University Press, 2010.
- [39] I. T. Voyles, C. J. Roy, Evaluation of model validation techniques in the presence of aleatory and epistemic input uncertainties, in: 17th AIAA Non-Deterministic Approaches Conference, ARC, 2015, pp. 1–16. doi:10.2514/6.2015-1374.
- [40] D. V. Lindley, On a measure of the information provided by an experiment, *The Annals of Mathematical Statistics* (1956) 986–1005.
- [41] J. Brynjarsdóttir, A. O’Hagan, Learning about physical parameters: The importance of model discrepancy, *Inverse Problems* 30 (11) (2014) 1–24.